

A MAN NEEDS A FEMALE LIKE A FISH NEEDS A LOBOTOMY: THE ROLE OF  
ADJECTIVAL NOMINALIZATION IN PEJORATIVE MEANING

Melissa Robinson

Thesis Prepared for the Degree of  
MASTER OF ARTS

UNIVERSITY OF NORTH TEXAS

May 2018

APPROVED:

Patricia Cukor-Avila, Major Professor  
Alexis Palmer, Committee Member  
Rosalia Dutra, Committee Member  
Sadaf Munshi, Interim Chair of the Department  
of Linguistics  
Kinshuk, Dean of the College of Information  
Victor Prybutok, Dean of the Toulouse Graduate  
School

Robinson, Melissa. *A Man Needs a Female like a Fish Needs a Lobotomy: The Role of Adjectival Nominalization in Pejorative Meaning*. Master of Arts (Linguistics), May 2018, 107 pp., 14 tables, 17 figures, references, 39 titles.

This thesis documents the grammatical processes and semantic impact of innovative ways to pejoratively reference individuals through adjectival nominalization. Research on nominalized adjectives suggests that when meanings shift from having one property (1) to becoming a kind with associated properties (2), the noun form often encodes stereotypical attributes: [1] “Her hair is blonde.” (hair color); [2] “He married a blonde.” (female, sexy, dumb). Likewise, the linguistic phenomenon of genericity refers to classes or kinds and different grammatical structures reflect properties in different ways. In 1 and 2 above, the shift from adjectival blonde to indefinite NP a blonde moves the focus from the definitional characteristic to the prototypical. Similarly, adjectival gay [3] is definitional, but the marked, nominal form [4] adds socially-based conceptions of the “average” gay (example from Twitter): [3] *jesus christ i make a joke and now im a gay man?* (sexuality) [constructed]; [4] *jesus christ i make a joke and now im a gay? ...* (flamboyant, abnormal). To investigate innovative reference via nominalization, two corpus studies based in human judgment were conducted. In the first study, a subset of the corpus (N=121) was annotated for pejoration by five additional linguists following the same guidelines as the original annotator. In the second study, 800 instances were annotated by non-experts using crowd-sourcing. In both studies we find a correspondence between nominal status and pejorative meaning.

Copyright 2018

by

Melissa Robinson

## ACKNOWLEDGEMENTS

I would like to thank my committee for their part in the development of this thesis. Each member had a large role this research. I am deeply grateful for Patricia Cukor-Avila's continued support through this project, for keeping me on track, reminding me of the important things, and encouraging me to present my research. Alexis Palmer, thank you so much for your constant and greatly needed advice. Thank you for introducing me to new methods of analysis and for your patience during my learning process. I am tremendously grateful to Rosalia Dutra for the encouragement and insight she has provided me from the very start, as it was your class that I first started exploring this topic.

Along with my committee I would like to thank Haj Ross for his help in developing the groundwork for this thesis. Exploring this topic as a squib during office hours provided me with much of the foundation that is my thesis today.

I would like to thank the support and encouragement from all my friends, including Jessica Protasio, Jessalyn Goodman, Darrin Miller, Mary Burke, Sumshot Khular, and many more. Additionally, thank you to Kristy Phillips for her contribution to the corpus and sentiment analysis studies.

Lastly, I am eternally grateful to my family for their understanding and support during a trying time. Without your belief in me, this thesis would never have come to be.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS .....	iii
LIST OF TABLES .....	vii
LIST OF FIGURES .....	viii
CHAPTER 1. INTRODUCTION .....	1
1.1 Statement of Problem.....	1
1.2 Hypotheses .....	7
1.3 Structure of Thesis .....	8
CHAPTER 2. CORPUS.....	10
2.1 Data .....	10
2.1.1 <i>Female</i> .....	11
2.1.2 <i>Illegal</i> .....	13
2.1.3 <i>Gay</i> .....	14
2.1.4 <i>Poor</i> .....	15
2.2 Data Extraction .....	17
2.3 Additional Grammatical Categories .....	17
2.4 Pejorative Annotation.....	18
2.5 Special Cases .....	20
2.6 Summary of the Corpus .....	22
CHAPTER 3. MULTIFACETED FEATURES OF PEJORATIVE NOMINALIZATION .....	23
3.1 Introduction.....	23
3.1.1 Cluster of Linguistic Properties .....	26
3.2 The Meaning of the Word .....	28
3.3 Contrasting in Pairs.....	31
3.4 Nominalization.....	35
3.4.1 The Zero Plural, Nominalization, and Adjectives with 'Empty' Nouns ....	35
3.4.2 Nominalization.....	38
3.4.3 Adjectives and Nouns .....	40
3.5 Reference .....	43

3.5.1	Specific Reference .....	43
3.5.2	Genericity: Kind-Referring .....	45
3.6	Prototypicality .....	48
3.7	Markedness .....	51
3.8	Grammatical Form and Associative Content .....	52
3.9	Conclusion .....	56
CHAPTER 4. A CLOSER LOOK AT THE DATA.....		57
4.1	Introduction.....	57
4.2	The Pejorative Noun <i>Female</i> .....	57
4.3	The Pejorative Noun <i>Poor</i> .....	63
4.4	The Pejorative Noun <i>Gay</i> .....	64
4.5	The Pejorative Noun <i>Illegal</i> .....	68
4.6	Conclusion .....	72
CHAPTER 5. EMPIRICAL STUDY: HUMAN JUDGMENT .....		75
5.1	Introduction.....	75
5.2	Expert Judgment Study .....	75
5.2.1	Results.....	77
5.3	Crowd-Sourced Non-Expert Annotation.....	78
5.3.1	Methods.....	79
5.3.2	Results.....	82
5.3.3	Non-Expert Annotation against Original Expert Annotation.....	83
5.4	Conclusions.....	84
CHAPTER 6. SENTIMENT ANALYSIS.....		85
6.1	Introduction.....	85
6.2	Methods.....	86
6.3	Data .....	87
6.4	Lexicons .....	88
6.4.1	MPQA Subjectivity Lexicon.....	89
6.4.2	Harvard General Inquirer .....	90
6.4.3	Bing Liu Opinion Lexicon .....	90
6.5	Scoring Methods and Evaluation .....	90

6.6	Results.....	91
6.7	Discussion .....	95
6.8	Conclusion .....	98
CHAPTER 7. CONCLUSION.....		100
7.1	Discussion and Implications .....	100
7.2	Future Research .....	100
APPENDIX: GRAMMATICAL FORMS .....		102
REFERENCES .....		104

## LIST OF TABLES

	Page
Table 2.1: <i>Female, Illegal, Poor, and Gay</i> Grammatical Forms .....	10
Table 2.2: <i>Female</i> Grammatical Forms .....	13
Table 2.3: <i>Illegal</i> Grammatical Forms .....	14
Table 2.4: <i>Gay</i> Grammatical Forms .....	15
Table 2.5: <i>Poor</i> Grammatical Forms .....	16
Table 5.1: Fleiss' Kappa Agreement between Expert Annotators .....	77
Table 5.2: Agreement between Annotators via Majority Vote .....	82
Table 5.3: Correlation between Grammatical Form and Sentiment .....	83
Table 5.4: Crowd-Sourced Annotations Compared against Original Annotator.....	84
Table 6.1: Sentiment Category Breakdown of <i>Female, Gay, Poor, and Illegal</i> .....	87
Table 6.2: Sentiment Analysis Results for <i>Female</i> .....	92
Table 6.3: Whole Instance Sentiment Analysis Results for <i>Gay</i> .....	92
Table 6.4: Whole Instance Sentiment Analysis Results for <i>Poor</i> .....	93
Table 6.5: Whole Instance Sentiment Analysis Results for <i>Illegal</i> .....	94



## LIST OF FIGURES

	Page
Figure 3.1: Paradigm of Nominal Forms for <i>Cat</i> .....	23
Figure 3.2: Paradigm of Nominal Forms for <i>Female</i> .....	24
Figure 3.3: Paradigm of Nominal Forms for <i>Gay</i> .....	24
Figure 3.4: Paradigms of Nominal Forms for <i>Illegal</i> and <i>Poor</i> .....	25
Figure 3.5: Creation of a Kind During Nominalization: <i>Blonde</i> (Wierzbicka) .....	38
Figure 3.6: Proposed Kind Creation for <i>Illegal</i> .....	39
Figure 3.7: Reference Spectrum .....	44
Figure 3.8: Acton’s Definite Plural Nonmembership .....	54
Figure 4.1: Contrastive Pairings of Pejorative <i>Female</i> .....	58
Figure 6.1: Sentiment Analysis Sentence Example .....	87
Figure 6.2: Sentiment Analysis Neutral Rating .....	87
Figure 6.3: Examples of Sentiment Scoring Over Lexicons .....	95
Figure 6.4: Whole Rating Example.....	96
Figure 6.5: Sentiment Analysis Isolated Clause Approach.....	96
Figure 6.6: Sentiment Analysis Positive Adjective.....	97
Figure 6.7: Sentiment Analysis Three Approaches.....	97
Figure 6.8: Sentiment Analysis False Neutral.....	98

## CHAPTER 1

### INTRODUCTION

#### 1.1 Statement of Problem

Within current discourse, there are a vast number of ways to imply pejorative meaning. Often the preferred method is to use overt, established derogatory language such as *bitch*, *wetback*, or *towel head*. Usually, these derogatory reference terms focus on a specific characteristic about a group and portray that characteristic in a blatantly negative light. These characteristics are often related to race/ethnicity (*beaner*), gender (*cunt*), or sexuality (*faggot*). While these explicit terms are typically what come to mind when discussing derogatory language, speakers also use subtler means of disparaging groups or individuals.

Microaggressions are another way to disparage individuals in a subtler manner than using blatant derogatory terms like the ones discussed above. Sue (2010, p. xvi) defines microaggressions as “brief, everyday exchanges that send denigrating messages to certain individuals because of their group membership.” The focus on group membership is very similar to the more deliberate terms mentioned above, the difference, however, is that microaggressions are usually viewed as innocuous as they often occur in casual, everyday exchanges. The power behind microaggressions is that a seemingly innocent comment implies and taps into stereotypical ideas about the individual and group they belong to.

Consider the following scenario:

[1] A man sees a woman walking down the street and smiles at her. Man: “You would be much prettier if you smiled more.”

The scenario in (1) is an example of a microaggression toward women. This kind of comment is often considered harmless, as nothing overtly negative was said; in fact, it includes positive words such as *pretty* and *smile*. However, this comment actually shifts the power

dynamic between the man and woman to imply some negative ideas about women in general listed below:

1. The man is making a judgment against the woman because she is not smiling.
2. He believes he has the right and authority to make and voice this judgment to the woman.
3. The man places value in the woman's appearance over all else.
4. The term *pretty* has connotations of cuteness and daintiness associated with women.
5. The man assumes being pretty is something the woman values.
6. The man feels he has the social space to make such a comment.

Sue (2010, pp. xvi-xvii) explains that exchanges like (1) “are so pervasive and automatic in daily conversations and interactions that they are often dismissed and glossed over as being innocent and innocuous”. As a result, recipients of microaggression often feel confused and uncertain how to react, or if they even have a right to be offended. While terms like *bitch* and *whore* are so overtly negative and difficult to ignore in everyday interactions, microaggressions function so subtly that often recipients will seem unreasonable if they do react.

Similarly, positive or neutral words can take on negative implications within certain contexts. Alim and Smitherman (2012) discuss this type of microaggression in commentary made about President Obama during the presidential campaign in 2008. Obama was often described in the press as *articulate* (as shown in Example 2), even by his future running mate Senator Joe Biden.

[2] He’s the first mainstream African American who is articulate and bright and clean and a nice-looking guy.  
(*New York Observer*, 2007)

While this comment has many positive words, Alim and Smitherman argue that the overarching sentiment of the comment has negative racial implications. Because *articulate* is

used to reference Obama, a Black candidate, the implication is that Black men are not articulate.

Moreover, Alim and Smitherson (p. 39) note:

If one needs to consistently point out that an individual Black person is “good,” “clean,” “bright,” “nice-looking,” “handsome,” “calm,” and “crisp,” it suggests that White private opinions about Blacks, in general, hold that they are usually the opposite— “bad,” “dirty,” “dumb,” “mean-looking,” “ugly,” “angry,” and “rough.” So, it’s not merely the use of *articulate* that’s problematic, nor the expression of the surprise or bewilderment that makes it suspect, it is also the fact that its adjectival neighbors describe qualities that help create these exceptionalizing discourses.

In other words, the negative meaning associated with *articulate* is amplified by pairing it with positive adjectives such as *clean* and *bright*. These positive adjectives exaggerate the ‘surprising’ quality that was initially implied with the use of *articulate*.

Much like microaggression and the contextual meaning of *articulate* described above, small shifts in language, such as using an adjective as a noun, can result in a shift in meaning as well. In this thesis, I will argue that certain adjectives can take on a negative meaning when they are used as nouns.

My interest in this topic results from having noticed that the word *female* was being used to refer to women in certain contexts. While I was exploring this use of *female*, the 2015 Presidential primaries started and soon another adjective that was used as a noun caused a negative reaction on social media. During the first Republican debate, the noun *illegal* was used quite heavily, including a memorable quote from Mike Huckabee:

[3] because the money paid in consumption is paid by everybody, including illegals, prostitutes, pimps, drug dealers, all the people that are freeloading off the system. (GOP Debate, Aug. 6th, 2015)

Following this remark, many comments were posted on Twitter criticizing the use of the word *illegals*. The following are some examples of the reactions:

- [4] “Illegals” is an absurd, offensive term. I’m an immigrant, a human. I shouldn’t be reduced to “an illegal,” Nor should others. #GOPDebate (Twitter, Aug. 6, 2015)
- [5] Reginald is all, “Did you really just refer to human people as ‘illegals’ again, really” #GOPDebate (Twitter, Aug. 6, 2015)
- [6] FETUS IS PEOPLE. BABIES IS PEOPLE. GAYS BLACKS ILLEGALS TRANS NOT PEOPLE. #GOPDebate (Twitter, Aug. 6, 2015)
- [7] Since when was “illegals” a noun? Merriam-Webster? #GOPDebate (Twitter, Aug. 6, 2015)

Examples 4-7 are only a small portion of the tweets reacting to this use of *illegal* in the August 6 GOP debates.

*Illegal* as a noun lacks the human quality that *immigrant* has. As seen in Examples 4-7, people have noticed it has negative shift in meaning as well. It became apparent to me that there were other adjectives, like *gay* and *poor*, that were also used in this way as count nouns.

Consider the following pairs of examples:

- [8] (a) i hate that females always scream RAPE when they clearly was getting a train ran on them and everyone found out (Twitter, April 14, 2016)
- (b) i hate that women always scream RAPE when they clearly was getting a train ran on them and everyone found out (Constructed)
- [9] (a) jesus christ I make a joke and now im a gay? Is that how you catch the gay? (Twitter, Sept. 18, 2016)
- (b) jesus christ I make a joke and now im a gay man? Is that how you catch the gay? (Constructed)
- [10] (a) Sanctuary cities should be ‘demolished’ f\*\*\* the illegals. (Twitter, Oct. 6, 2016)
- (b) Sanctuary cities should be ‘demolished’ f\*\*\* the immigrants. (Constructed)

[11] (a) “if you receive your tax return and you dont need it, shred it or else a poor will get it by accident”. (Twitter, Jan 14, 2016, sarcastic read)

(b) “if you receive your tax return and you dont need it, shred it or else a poor man will get it by accident”.  
(Constructed)

In sentences 8-11 (a), the underlined phrases are examples of common adjectives used as nouns. In the constructed 8-11 (b) sentences, the forms are replaced with an adjective form modifying a noun, such as *a poor man* or alternative noun such as *woman*. When comparing underlined forms in 8-11 (a) with the underlined phrases in 8-11 (b), a subtle pejorative meaning is revealed. While 8-11 (b) might still be perceived as negative, this perception has more to do with the overall context of the sentences rather than the underlined forms. This thesis argues that the pejoration seen in 8-11 (a) derives from the adjective to noun shift rather than the overall negative context within the sentence.

I use the term *innovative* to describe these noun forms of *illegal*, *female*, *gay*, and *poor*; however, the pejorative use of these nouns is not a recent change. The Oxford English Dictionary lists a depreciative use of the noun *female*, describing it as “a generic descriptor implying low class or lack of traditional feminine qualities” (Female, n. and adj., 2017). Examples for this form date back to 1425, though the following two examples dating from 1849 and 1889 respectively appear more closely align with the type of pejorative meaning studied in this thesis:

[12] “The ‘Totty’ of the present day: and his female, (for the creature can scarcely be dignified by the name of woman).” (Female, n. and adj., 2017)

[13] “There are no ladies. The only word good enough for them is the word of opprobrium—females.” (Female, n. and adj., 2017)

The OED also has citations of *illegal* as a noun to reference immigrants that date back to 1939, *gay* as a noun since 1953, and *poor* as a singular and plural noun, though now rare or archaic, since the 13<sup>th</sup> century. Although the use of these terms has been recorded, in this thesis I will show that they have been given new life and in some cases, have become a slightly more pejorative in meaning.

The process of adjectives used as nouns, which I refer to as adjectival nominalization<sup>1</sup>, is a productive process in English. In other words, speakers are familiar with the process of converting adjectives to nouns, and therefore they are also able to create various innovative forms. Consider the following Twitter discussion:

[14] **S1:** I'm collecting adjectives used as nouns in common language:

A lame

A grown

An old

A poor

Is this a thing now? @OxfordWords @Dictionarycom

**S1:** Maybe @GrammarGirl or @AllusionistShow knows? Am I just a confused or is this a thing?

**S1:** Context: the sentence “he's a lame” was used in a TV show.

**S1:** I've heard people use the term “a grown” and the plural “grows” to describe adults multiple times.

**S1:** A student said to me “oh, you're an old” upon learning my age.

**S1:** I've seen the term “a poor” and plurals “poors” used on social media by several different groups.

**S1:** I am perplexed.

**S2:** Never heard any of those but then again I don't talk to people.

(Twitter, May 8, 2017)

---

<sup>1</sup> For my purposes, adjectival nominalization is defined as adjectives used as nouns that can take the plural marker -s as well as possessive -'s. I discuss this further in chapter 3.

S1 has noticed that adjectives are being used as nouns and has offered up a list of examples, *a lame*, *a grown*, *an old*, and *a poor*, as evidence of this trend.

While this process of adjectives used as nouns occurs with numerous adjectives, the present study limits its scope to four adjectival nominals, *poor*, *illegal*, *gay*, and *female*. The focus of the study is not on the terms themselves, but on the pejorative process and the reasons behind the shift in meaning created when these four adjectives are nominalized.

## 1.2 Hypotheses

The purpose of this study has two parts. In the first part I will examine the various elements involved in the pejoration process, such as nominalization, genericity, semantic meaning in the grammatical forms<sup>2</sup>, and associative content, to explore why some adjectives, e.g. *legal*, *straight*, *rich*, and *male*, can be used as nouns yet do not have pejorative meaning. Next, I will discuss two empirical approaches, human judgment and computational methods, used to analyze subtle semantic shifts that occur in the pejoration process.

My hypotheses are as follows:

- The pejoration in *illegal*, *female*, *gay*, and *poor* is not a result of one element, such as nominalization, but a cluster of properties that provides the appropriate environment for pejoration.
- Adjectives can soften a negative or stigmatized property, whereas nouns amplify a negative or stigmatized aspect. Therefore, speakers favor adjectives when their intent is to soften a stigmatized property, and they favor nouns when the intent is to amplify the negative or

---

<sup>2</sup> In this thesis, *grammatical forms* refers to the forms the nouns can occur in, specifically indefinite singular, definite singular, bare plural, and definite plural.



stigmatized property. An example of this is the use of an adjective in the phrase *a crippled man* rather than the noun phrase *a cripple*, which is perceived as more negative.

- Some adjectives that are nominalized do not shift to a negative meaning. In order to be perceived as pejorative, an adjective requires several aspects, including markedness<sup>3</sup>, social stigmatization, and negative association with the term.

### 1.3 Structure of Thesis

In Chapter 1, I introduce the topic of study, which is the negative shift in meaning that occurs when *illegal*, *poor*, *gay*, and *female* are used as nouns. The remainder of the thesis is divided into 6 chapters. Chapter 2 introduces the data and its sources, and the methods used for coding and analysis.

Chapter 3 discusses the semantic and pragmatic elements associated with the shift in meaning as a result of the nominalization process. This chapter presents the various linguistic properties that leads to the negative meaning, with each property discussed in detail. These properties include markedness, nominalization, reference, and prototypicality.

Chapter 4 presents a closer look at the data and discusses in detail the unique aspects of each of the four nominalized forms under study.

In Chapters 5 and 6, I present empirical and computational approaches to the corpus. In Chapter 5, I discuss the results from empirical studies using human judgement of the target forms. Chapter 6 presents a sentiment analysis study, where computational approaches are applied to analyze the subtle semantic shifts in *illegal*, *gay*, *female*, and *poor*.

---

<sup>3</sup> Markedness is used in this thesis to reference the unexpected or unusual grammatical forms that would likely be perceived as ungrammatical. This is in comparison to the forms that are fully accepted as grammatical.

Lastly, in Chapter 7, I summarize what has been revealed by this study, and additionally, I discuss its implications and possible areas for future research.

## CHAPTER 2

### CORPUS

#### 2.1 Data

To investigate the hypotheses outlined in Chapter 1, I needed to collect examples of *illegal*, *gay*, *poor*, and *female* being used as nouns to analyze how and in what context these forms were being used. In this chapter, I describe the methods of my data collection, the coding of the various grammatical forms<sup>4</sup>, as well as the varied sources where these forms were found. Instances of *female*, *illegal*, *poor*, and *gay* used as nouns were collected and categorized as various nominal constructions. The overall goal was to annotate<sup>5</sup> each example as pejorative and non-pejorative, so guidelines were written for annotators to distinguish between pejorative and non-pejorative uses. Table 2.1 presents the totals of each nominal construction:

Table 2.1: *Female, Illegal, Poor, and Gay* Grammatical Forms

	Examples	<i>Female</i>	<i>Illegal</i>	<i>Poor</i>	<i>Gay</i>
Indefinite singular	<i>A poor</i>	122	105	35	12
Definite singular	<i>The gay</i>	30	8		3
Bare plural	<i>Females</i>	401	238	73	80
Definite plural	<i>The illegals</i>	62	157	203	41
Quantified singular	<i>Any poor</i>	19		1	2
Quantified plural	<i>Many gays</i>	45	28	4	6
Demonstrative singular	<i>This female</i>	7	1	1	
Demonstrative plural	<i>Those illegals</i>	20	6	9	2
Pronoun singular	<i>My poor</i>	2			
Pronoun plural	<i>His females</i>	2		5	
Vocative plural	<i>You gays</i>	7		5	
Total		717	543	336	146

---

<sup>4</sup> In this thesis, *grammatical form* refers to the different nominal constructions, such as indefinite singular, definite singular, bare plural, and the definite plural.

<sup>5</sup> In this thesis, the terms *coding* and *annotating* are used more or less interchangeably. However, coding is often used when coding specific linguistic features, such as grammatical form. Annotating is used to mark aspects that have some room for interpretation, such as annotating for sentiment.

### 2.1.1 *Female*

The data for *female* was collected from both written and spoken sources. The majority of the examples were extracted from commentary found on various Men's Rights forums, blogs, and video blogs. The Men's Rights movement is a collection of various online groups that claim to promote rights needed by men. However, it is common for the discussion in these forums and blogs to focus on anti-feminism topics rather than men's rights. One such group is MGTOW (Men Going Their Own Way) which is a type of men's right group. Bloggers who are MGTOW followers often include the term MGTOW either in the title or as a hashtag. This identifying factor was used as a search tool to find various blogs, video blogs, and forums that were tied to this ideology.

Additional data on *female* comes from a transcribed CNN interview on the topic of whether a woman should be president. During this interview both the reporter and the woman being interviewed switched between *woman* and *female*, which provided examples in context of how speakers shift between these nouns. Instances of *female* were also found on Twitter and Reddit posts.

*Female* can be used for biologically based discussions or to include children as well as adults within the same reference, so only the use of *female* in place of *woman* was considered when collecting data. When referring to a human, *woman* would be the more standard<sup>6</sup> term because the semantic properties of *woman* are +FEMALE, +ADULT, +HUMAN, whereas *female* only encodes biological sex. The nominalized *female* forms analyzed in this study are

---

<sup>6</sup> The term *standard* is used in this thesis to describe more widely used, readily accepted grammatical forms and language uses. My intended use of *standard* is not to make a judgment or distinguish *good language* from *bad language*, but rather to differentiate widely used language from more innovative language for the sake of discussion.

restricted to uses where the conventional noun *woman* would seem more appropriate, such as

Example 15:

[15] Real women don't like being compared to the basic bitches you were with before, if they hurt you, they hurt you, understand not all women are going to do that to you. But I digress, back to you weak ass females, you had a good man but you treated him like shit. (YouTube.com, March 4, 2015)

The use of *females* in Example 15 is not biologically gender-related, nor is it referencing a wider age range, to include children, that would require the use of the more general term *females*.

Instead, Example 15 reveals a different motivation on the part of the speaker for using *female*, to imply negative meaning. Many of the examples collected reveal this distinct differentiation between *women* and *females*. Differences in the use of *female* were taken into consideration when collecting the data in an attempt to restrict the tokens to true pejorative uses.

Several coding constraints apply to the tokens of *female*. First, I only include cases where the referent of *female* is human – instances of *female* used to reference animals were not included. Second, I considered the purpose of the use of *female*. There are settings where *female* is more appropriate than *woman*, such as its use in law enforcement (Example 16) and medical and scientific discourse or when referencing across a larger age group to include children or teenagers.

[16] The suspect is a female of unknown age, proceed with caution. (Constructed)

Additionally, there are times when the topic of discussion is biologically gender focused, such as a discussion on the biological and instinctual differences between males and females, seen in Example 17.

[17] Estrogen makes females more emotionally driven on average compared to males. (YouTube commentary, 2015)

I included these forms in the corpus and coded them separately from the forms under study.

There is a total of 717 examples of *female* included in the corpus. One hundred twenty-two tokens are indefinite singulars, 30 definite singulars, 401 bare plurals, 62 definite plurals, 19 quantified singulars, 45 quantified plurals, 7 demonstrative singulars, 20 demonstrative plurals, 2 pronoun singulars, 2 pronoun plurals, and 7 vocative plurals.

Table 2.2: *Female* Grammatical Forms

<i>Female</i>	<i>Examples</i>	#
Indefinite singular	<i>A female</i>	122
Definite singular	<i>The female</i>	30
Bare plural	<i>Females</i>	401
Definite plural	<i>The females</i>	62
Quantified singular	<i>Any female</i>	19
Quantified plural	<i>Many females</i>	45
Demonstrative singular	<i>This female</i>	7
Demonstrative plural	<i>Those females</i>	20
Pronoun singular	<i>My female</i>	2
Pronoun plural	<i>Their females</i>	2
Vocative plural	<i>You females</i>	7
Total		717

### 2.1.2 *Illegal*

The data collected for *illegal* was primarily found in political discourse, both formal and informal. Public discussions of the 2016 Presidential debates and the upcoming election were one source of instances of *illegal*. Additionally, a small portion of the data was collected from news interviews with political candidates, but the majority of the data was collected from commentary on Reddit, Twitter, and forums that were often labeled as Alt Right. The use of *illegal* as a noun was strongly linked to topic. For example, many instances came from posts about immigration and building a wall between the US and Mexico.

It was very rare to find *illegal* used to reference anything but an immigrant. There were three instances within the corpus that were coded separately to distinguish them from the forms in question. Two were human referents, one referencing under-aged girls and one referencing someone under the drinking age. The third example was the non-human referent of *illegal fireworks*.

There are 543 instances of *illegal*: 105 indefinite singulars, 8 definite singulars, 238 bare plurals, 157 definite plurals, 28 quantified plurals, 1 demonstrative singular, and 6 demonstrative plurals.

Table 2.3: *Illegal* Grammatical Forms

<i>illegal</i>	Examples	#
Indefinite singular	<i>An illegal</i>	105
Definite singular	<i>The illegal</i>	8
Bare plural	<i>Illegals</i>	238
Definite plural	<i>The illegals</i>	157
Quantified plural	<i>Many illegals</i>	28
Demonstrative singular	<i>This illegal</i>	1
Demonstrative plural	<i>Those illegals</i>	6
Total		543

### 2.1.3 *Gay*

The data for *gay* was primarily collected on Twitter and Reddit. This term presents a unique challenge as speakers within the gay community can use nominal forms of *gay* without them having a pejorative meaning. However, if someone outside the gay community uses *gay* in the same way it can be perceived as pejorative in meaning, as seen in Example 18.

[18] Only the gays or allies of the gays get to call us “gays”, @realDonaldTrump. “Gay people” would be respectful term from your type. #debate (Twitter, Oct. 19, 2016)

Note that in this tweet, the interlocutor expresses that whether someone is in the gay community

or not makes a big difference in how the nominalized form of *gay* will be perceived. Here, Donald Trump is not viewed as a member of the gay community and therefore his use of *gay* is felt to be offensive and he is directed by the poster to use the adjectival form instead.

This restricted use within the gay community is not unique to the use of *gay*. It is quite common for derogatory terms, such as *faggot*, *queer*, or *dyke*, to be used in the gay community with more neutral rather than pejorative meaning as an effort to regain power by reclaiming a negative term through reappropriation (Galinsky et al., 2013). Galinsky et al. explain that in self-labeling with a normally derogatory term, the negative meaning softens and is sometimes stripped of negativity by redefining it through positive self-use. However, these terms still have pejorative meaning when used outside the community.

Table 2.4 shows a breakdown of *gay* by grammatical form. There is a total of 146 instances of *gay* in the corpus: 12 indefinite singulars, 3 definite singulars, 80 bare plurals, 41 definite plurals, 2 quantified singulars, 6 quantified plurals, and 2 demonstrative plurals.

Table 2.4: *Gay* Grammatical Forms

<i>Gay</i>	Examples	#
Indefinite singular	<i>A gay</i>	12
Definite singular	<i>The gay</i>	3
Bare plural	<i>Gays</i>	80
Definite plural	<i>The gays</i>	41
Quantified singular	<i>Any gay</i>	2
Quantified plural	<i>Many gays</i>	6
Demonstrative plural	<i>Those gays</i>	2
Total		146

#### 2.1.4 *Poor*

Data for *poor* was collected from Twitter, Reddit, and commentary on various informal



news sites, as well as a formal political column from *The New York Times*. While collecting the data, it became apparent that nominalized *poor* is often, but not always, used in a sarcastic or satirical manner. More specifically it is used to voice another's perceived attitude in a mocking way. Example 19 shows this use, as the poster is not Donald Trump, but yet is an imitation of what the speaker perceives to be Trump's attitude toward the poor:

[19] Trump: "We have a divided nation." And I plan to keep it that way. Keep it in the inner city, you poors. #debate (Twitter, Oct. 9, 2016)

The use of *poor* in Example 19 is meant to exaggerate the perceived attitude of Trump. This use suggests that the speaker is aware of the pejorative meaning in the nominalized form of *poor*.

Instances like 19 were coded as sarcastic/satirical.

Table 2.5 shows a breakdown of *poor* by grammatical form. There are 336 instances: 35 indefinite singulars, 73 bare plurals, 203 definite plurals, 1 quantified singulars, four quantified plurals, 1 demonstrative singular, 9 demonstrative plurals, 5 pronoun plurals, and 5 vocative plurals.

Table 2.5: Poor Grammatical Forms

<i>Poor</i>	Examples	#
Indefinite singular	<i>A poor</i>	35
Definite singular	<i>The poor</i>	0
Bare plural	<i>Poors</i>	73
Definite plural	<i>The poors</i>	203
Quantified singular	<i>Any poor</i>	1
Quantified plural	<i>Many poor</i>	4
Demonstrative singular	<i>That poor</i>	1
Demonstrative plural	<i>Those poors</i>	9
Pronoun plural	<i>His poors</i>	5
Vocative plural	<i>You poors</i>	5
Total		336

## 2.2 Data Extraction

The data for the corpus was manually extracted. A manual approach allowed me to carefully verify the forms in question and filter out problematic forms, such as *pours* spelled as *poors* which was a common mistake. Another advantage was the ability to determine the referents of the four terms. This is especially important for *female*, which is often used to reference animals.

The extraction was done by using various search methods for the target forms. Twitter's advanced search method allowed me to restrict the search to a specific grammatical form, such as *the poors*. A similar method was used to extract data from YouTube. Search engines, such as Google, were used to find written blog sources and articles that contained uses of the target forms. Lastly, searches were conducted based on topics that would likely produce high levels of the target forms. For example, the site Reddit is broken down by subreddits that are loosely based on topic. Subreddits such as *The\_Donald*, which is a forum for Donald Trump supporters, had various threads on the topic of immigration. Once a thread was located, I searched the conversation for the target forms.

Using search engines worked well for finding plural forms. The singular forms, however, presented more of a challenge, as often the search results returned the adjective forms rather than the singular noun forms. To resolve this issue common verbs, such as forms of copular *be*, were added after the singular form while searching, resulting in instances like Example 20:

[20] Yeah dude being poor happens from time to time, but being A poor is a way of life. LOL (Twitter, Jul. 13, 2016)

## 2.3 Additional Grammatical Categories

Each example was coded for grammatical form as well as for pejoration. The main

categories for grammatical forms are indefinite singular, definite singular, bare plural, and definite plural. The corpus also includes demonstrative, number, quantifier, and personal pronoun forms as separate categories because these forms could influence whether an article is used or not used. For example, with a form that uses a demonstrative such as *those illegals*, the definite article will not be an option. Since this form cannot be considered a definite plural or a bare plural, and as grammatical form is relevant to the analysis, these forms were categorized separately. Definite plurals modified with a relative clause are also in a separate category, as in Example 21:

[21] Do the #illegals who were given greencards supposedly by accident factor into #HRC vetting #debates #Trumptrain  
(Twitter, Oct. 2016)

## 2.4 Pejorative Annotation

The main objective of this study was to annotate the instances for pejorative or non-pejorative meaning. The following categories were used to code for pejoration: *pejorative*, *non-pejorative*, *satire*, and *uncertain*. The *uncertain* category was used with instances where the sentiment is unclear or if something else makes the instance or target form itself questionable. The following guidelines were written for deciding if an instance is pejorative or non-pejorative. If the instance follows one or more of these guidelines the annotator codes the instance as pejorative:

- Negative adjective(s) modifying the target nominal form (22)
- Co-occurrence with phrases referring to particular stereotypes or behaviors associated with the relevant referent group (e.g. *freeloading* with an occurrence of *poor*) (23)
- Appearance near negative verbs such as *hate* or *despise*, or negative phrases such as *get rid of* or *hardly any good* (24)
- Co-reference with other negative terms, such as *slut* for *female* or *wetback* for *illegal* (25)

- Other negative implications not tied to a specific lexical item or phrase. (26)
- [22] “You have the distinct odor of poverty. Trust me, I can smell you from here! Sad filthy poors.” -Trump in PA (Twitter, Oct. 10, 2016)
- [23] Why don't gays like being girly? Cause a gay is normally called girly. (Twitter, Aug. 13, 2016)
- [24] Whites hate illegals. Blacks hate illegals. Native Americans hate illegals. Asians hate illegals. legals hate illegals. (Reddit, May 2016)
- [25] Hillary: Economic Plan: Keep the illegals coming, don't care if terrorists, we need their votes and too many dumb people vote for me anyway. (Twitter, Oct 5, 2016)
- [26] This feminist nonsense is to give every man the daily message that A Man Needs a Female Like a Fish Needs a Lobotomy. (YouTube commentary, 2016)

Guidelines for the *satire* category are not as prescribed because there is quite a bit of variation.

Some satirical forms are easily recognized as they are clearly voicing the perceived attitude of someone else. Many of these included labels of who this person is, such as Hillary Clinton (25) or Donald Trump (27).

- [27] Trump: “We have a divided nation.” And I plan to keep it that way. Keep it in the inner city, you poors. #debate (Twitter, Oct. 9, 2016)

However, there are other examples that are more complicated, for example when the speaker appears to be voicing a perceived attitude of a group or society as a whole but with no explicit declaration (28).

- [28] currently on a bus like some poor, I want to make it clear that I am not a poor, I am extremely wealthy, I just take buses to view the poors [loudly dictating into a tape recorder] Curiously, holding books and papers, it seems the poors are...pretending to be literate I heard a rumor poors could vote, like, in our elections (Twitter, Jul. 6, 2016)

Annotating examples like 28 mostly rely on context and the annotator's linguistic instinct rather than specific guidelines. Still, there is a distinct difference between Example 28 and an instance that would be coded as pejorative. The overall tone of 28 is sarcastic along with the ‘stage direction’ of *loudly dictating into a tape recorder*.

Sulis et al. (2016) argue that sarcasm usually involves positive sentiment words that imply a negative meaning within the sarcastic context. However, the sarcasm found in this corpus does not share the same characteristics as the sarcasm discussed by Sulis. Rather, it appears to have the unique feature of containing negative sentiment as seen in Example 29:

[29] hell NO the poors don't deserve solid bowel movments!  
(Wonkette.com, Apr. 27, 2015)

If sarcasm uses positive sentiment words to imply a negative sentiment, it stands to reason that the speaker in Example 29 actually is not expressing a personally held negative sentiment of *poors*, but rather is using the marked nominal form *poors* to make a statement of a perceived attitude of another. In this case, this was a comment in reaction to a satirical article advocating against Maine’s governor Paul LePage’s food restrictions on food stamps. More specifically, Example 29 is a comment on the article where readers were satirically debating what foods the poor should be allowed to eat as commentary on LePage’s legislation.

## 2.5 Special Cases

During the annotation process, some instances were judged to be too problematic for the pejorative analysis. One such situation is when the spelling and grammatical forms of the target forms were questionable, such as in Example 30.

[30] They are if a poorz has one or both. (Wonkette.com, April 27, 2016)

The target form in Example 30 has a conflicting grammatical form; while it has the indefinite article *a*, it also has an inventive spelling of the plural marker *-s*. In this situation, there is no clear way to determine if this is meant to be an indefinite singular or a plural, therefore it cannot be coded for grammatical form, so it is separated from the main corpus using the code *questionable form*.

Forms where the intended referent was different than the target referent, for example when *illegal* did not refer to *immigrant*, were coded as *different referent*. Example 31 and 32 illustrate this with two instances of *illegal* used for different human referents, one to reference underage girls (31) and another to reference people under the legal drinking age (32).

- [31] **S1:** @S2 lol i got one, for the people who dont i jus snap em like “who this”  
**S2:** @S1 lmao that's not necessary 🤔🤔🤔  
**S1:** @S2 lmao it is i needa kno who watchin my shit kause the jakes is hot  
**S2:** @S1 🤔🤔🤔 stop it.  
**S1:** @S2 🤔🤔🤔 frfr an i be havin the illegals  
**S2:** @S1 🤔🤔🤔 faceahh  
(Twitter, Oct. 6, 2016)

- [32] Dunno if this is still true, but used to be an ILLEGAL wasn't considered a man unless he could finish 18 pack and drive home. (Twitter, May 17, 2014)

Along with the unexpected human referent forms above, one instance of *illegal* to reference *illegal fireworks* is seen in Example 33.

- [33] An illegal went off on the ground and the sparks flew EVERYWHERE and one of them hit my forehead L00000L (Twitter, Jul. 4, 2015)

Finally, the target forms that were used to reference the term itself, as in Example 34, were coded as *term reference*.

[34] sorry, but calling someone an illegal isn't racist! Illegal isn't a race. (Twitter, Jun 26, 2016)

All of these special cases were coded separately to remove them from the analysis. In total there were 13 special cases for *female*, 22 for *illegal*, 6 for *poor*, and 6 for *gay*.

## 2.6 Summary of the Corpus

This chapter has covered the details of the data collection, and the coding and annotation methods used. The following is a bullet-point summary list of the corpus:

- Made up of four nominal target forms: *illegal*, *female*, *poor*, and *gay*
- Organized into four main grammatical forms: indefinite singular, definite singular, bare plural, and definite plural
- Additional grammatical forms: demonstratives, quantifiers, and personal pronouns.
- Sources included Twitter, Reddit, video and written blogs, and news articles
- Four categories for sentiment: pejorative, non-pejorative, satire/sarcasm, unclear

The next chapter discusses various semantic aspects related to *female*, *illegal*, *gay*, and *poor* and presents a hypothesis as to why these adjectives are perceived negatively when used as nouns.

## CHAPTER 3

### MULTIFACETED FEATURES OF PEJORATIVE NOMINALIZATION

#### 3.1 Introduction

In English, count nouns have four possible grammatical forms to indicate number. In Figure 3.1, there are four forms using the noun *cat*. This paradigm includes the indefinite singular form *a cat is*, the definite singular *the cat is*, the bare plural *cats are*, and the definite plural *the cats are*. The singular forms indicate one cat while the plural forms indicate more than one cat. Additionally, the presence or absence of articles tells us something about whether the noun has been referenced previously in the discourse and if the noun references something specifically or generally. The indefinite article in *a cat* signals that the speaker is not referencing a specific cat. In contrast, the definite article in both the singular and plural forms indicate that the speaker is referencing a specific cat or cats and has likely mentioned the reference earlier on in the discourse. In the bare plural form, there is no article to indicate definiteness or indefiniteness. Instead, the bare plural functions as a form that references a class or category, such as the category of *cats*.

<i>a cat is</i>	<i>cats are</i>
<i>the cat is</i>	<i>the cats are</i>

Figure 3.1: Paradigm of Nominal Forms for *Cat*

Likewise, *female* follows the same number paradigm as *cat*, as is shown in Figure 3.2. While *female* is often used as an adjective, it is also well established as a noun and can be used without restrictions in the paradigm.



<i>A female is</i>	<i>Females are</i>
<i>The female is</i>	<i>The females are</i>

Figure 3.2: Paradigm of Nominal Forms for *Female*

With *gay* in Figure 3.3, the paradigm begins to reveal some holes or restrictions, specifically in the indefinite singular and definite singular forms. In other words, while the bare plural and definite plural forms of *gay* are considered acceptable to standard grammar, the singular forms are not. Like *female*, *gay* is favored as an adjective but is established as a noun, although the range of nominal forms available for *gay* is more limited than for *female*. As seen in Figure 3.3, *gay* is established in the bare and definite plural forms, but not in the indefinite and definite singulars.

* <i>A gay is</i>	<i>Gays are</i>
* <i>The gay is</i>	<i>The gays are</i>

Figure 3.3: Paradigm of Nominal Forms for *Gay*

Due to these holes in the grammatical paradigm, a speaker would be required to revert back to the adjective form in order to indicate a singular referent, such as *a gay man is* or *the gay man is*.

The noun paradigm for count nouns is productive with native speakers of English. Therefore, the holes in the paradigm in Figure 3.3 can potentially be filled with innovative forms following the productive paradigm. The collection of indefinite and definite singular forms of *gay* within the corpus shows that speakers are filling these holes with innovative forms, following the productive noun paradigm as seen in Figures 3.1 and 3.2.

Figure 3.4 shows that *gay* is not the only form with holes in the paradigm. The paradigms for *illegal* and *poor* reveal further restrictions on grammatical forms:

<i>An illegal immigrant is</i>	<i>The illegal are</i>	<i>A poor man is</i>	<i>The poor are</i>
<i>*an illegal is</i>	<i>*illegals are</i>	<i>*a poor is</i>	<i>*poors are</i>
<i>*the illegal is</i>	<i>*the illegals are</i>	<i>*the poor is</i>	<i>*The poors are</i>

Figure 3.4: Paradigms of Nominal Forms for *Illegal* and *Poor*

In the first boxes in Figure 3.4, the adjective form is modifying a noun as shown with *an illegal immigrant* and *a poor man*. Additionally, the noun can be deleted and just the adjective would remain in its place, such as in *the illegal are* and *the poor are*. I refer to this form as the zero plural within this thesis as it has two distinguishing characteristics: there is no plural -s marker and it has plural agreement with the verb. The zero plural references the class or category of *illegal* or *poor* rather than specific individuals. This zero plural form is used with many adjectives, such as *the hungry are*, *the needy are*, and *the strong are*. However, this zero plural form cannot be used with nouns, such as with *cats*, *\*the cat are*. Further discussion of the zero plural can be found in Section 3.4.1.

In comparing the paradigm in Figure 3.1 to the paradigms for *illegal* and *poor* in Figure 3.4, it is apparent that both paradigms are entirely empty for *illegal* and *poor*. None of the noun forms are grammatical for either *illegal* or *poor*; instead a speaker following the standard norms would need to use the zero plural form. However, as discussed with the innovative singular forms of *gay*, speakers can fill in the holes of the paradigms. The difference between the innovative forms for *gay* and the innovative forms for *illegal* and *poor* is that *gay* has the plural forms established in the paradigm, but *illegal* and *poor* do not have corresponding forms in the paradigms. In using an innovative form for *illegal* or *poor*, the speaker is introducing an entirely new innovative set of forms. This would indicate that the innovation of these forms would be more jarring than the innovative singular forms of *gay*. Consider Examples 35-37 below:

[35] LOL So a Poor wants to run for President? What does Sanders know about the common American then. (Twitter, Apr. 15, 2016)

[36] JOSE THE ILLEGAL JUST WESTERN UNIONED IT TO MEXICO. (Reddit, 2016)

[37] Woke up to the news that Robert Pattinson came out as a gay? Is it true? Oh my Cedric Diggory! (Twitter, Sept. 20, 2016)

While the use of the indefinite singular of *gay* in Example 37 is still pejorative, the indefinite singular forms of *poor* (35) and *illegal* (36) feel slightly more marked as there is an expectation that a noun will follow. However, with the recent 2016 presidential election and the increase of political discourse on immigration, the nominal *illegal* is being used more and more. This will likely lead to a normalizing of the marked construction.

As speakers fill these holes within the paradigm with innovative forms, I find that these forms often are used with a pejorative meaning. However, in addition to these newly created forms, there is a cluster of linguistic properties that contributes to a pejorative meaning which is briefly outlined below.

### 3.1.1 Cluster of Linguistic Properties

- Meaning of the word (Section 3.2): Each adjective carries a small dose of negative meaning which leads to a pejorative meaning when they are used as nouns.
- Contrasting in pairs (Section 3.3): *Illegal*, *female*, *poor*, and *gay* are part of a pair, such as *rich* and *poor*, though the pejorative meaning varies when the adjectives are used as nouns.
- Filling the hole in the paradigm (Section 3.1): 3.1 As discussed above in Figures 3.3 and 3.4, the nominal forms in question are a result of speakers filling in the holes of the paradigm with innovative forms or creating a new paradigm where there wasn't one.

- Pejoration amplified by nominalization (Sections 3.5): The nominalization process, along with connected aspects listed below, amplifies the pejorative meaning.
  - Thingifying<sup>7</sup>/Dehumanizing (Section 3.2): When the adjective becomes the noun, a property now becomes a ‘thing’ and there is a dehumanizing factor to this process. The humanizing noun, such as *man* or *immigrant*, which include the semantic property +HUMAN, is replaced by an adjectival property that lacks this property.
    - (i) *An illegal immigrant* came into my shop today.
    - (ii) *An illegal* came into my shop today.
  - Stereotype (Section 3.4.2): When an adjective becomes a noun, the noun takes on stereotypical properties associated with the class or category of the referent.
    - (i) *Blonde* hair (Property: +Hair color)
    - (ii) *A blonde* (Stereotypical properties: +dumb, +sexy)
- Reference
  - Specific reference (Section 3.5.1): When innovative forms are used to reference specific entities, the pejorative meaning is more salient.
  - Kind referring generic reference (Section 3.5.2): When referencing a kind<sup>8</sup> (or class), generic statements can tie back to the stereotypes and prototypes now associated with the innovative nouns.
    - (i) *Poors* are a drain on the economy. (Constructed)

---

<sup>7</sup> The term *thingifying* references the act of making an adjectival property into a ‘thing’ when it is used as a noun.

<sup>8</sup> For my purpose, *kind* is used to reference the type or category of entity, such as *blondes* or *lions*. The term *kind* is often discussed with generic reference, but this thesis also references *kind* in relation to nominalization and prototypicality.

- Prototypicality (Section 3.6): This newly formed class or category is based on a prototypical idea of what type of entity belongs in this class. Stereotypes, which are derived from socio-cultural ideas, often lead to prototypes which is a cognitive feature.

- (i) The prototypical lion is an adult, has a mane, roars, etc.

- Markedness (Section 3.7): The innovative forms that are not provided in the paradigm seem strange or ‘marked’ as compared to the more standard grammatical forms.

- (i) *A poor man* was begging on the corner. (Unmarked)

- (ii) *A poor* was begging on the corner. (Marked)

In the following sections, these aspects are discussed in detail. While each is a separate facet, they all fit together to further develop the pejorative meaning.

### 3.2 The Meaning of the Word

The adjectival meanings of *illegal*, *female*, *poor*, and *gay* are important to starting point of the pejorative meaning. Bolinger (1980) argues that generally adjectives are words that carry bias, due to their elements of degrees, as well as biased pairs, such as *old/young*, *tall/short*, *fat/thin*. Additionally, most contrastive adjectival pairs function in relation to each other along a negative and positive scale. For example, with the pair *old/young*, *young* would lean toward the positive end of the scale while *old* would be more negative. Likewise, the adjectives *poor*, *gay*, *female*, and *illegal* are all on the negative end of the scale in comparison to their more positive adjective pairs, *rich*, *straight*, *male*, and *illegal*. This negative element at the adjectival level is part of why the innovative noun forms have a pejorative meaning.

For *poor* the definition is simply a lack of wealth, but there is a negative social class implication that goes beyond the monetary element as well. The concept of *poor* stretches

beyond lack of money to include lower education level and lack of etiquette. In comparison, *rich* has highly desirable qualities that go beyond wealth. The concept of *rich* includes perceptions of higher education and sophistication. This leaves *poor* on the negative end of the contrastive pair scale and *rich* on the positive end.

The basic meaning of *gay* is simply a sexual attraction to the same sex. While the negative bias might not be apparent in the meaning, the adjective has strong ties to negative use due to social perspectives on homosexuality. Additionally, in certain pairs, there is a social default, meaning one adjective in the pair is the *assumed* and one is the *marked*. *Straight* is the socially assumed sexuality whereas *gay* would be the marked sexuality. In other words, an individual is usually assumed straight until outside influence change this assumption. In this way, *straight* leans more toward the positive end of the spectrum with *gay* is more on the negative end.

*Female*, in the same way, is the marked adjective while *male* is the default. This marked sex can be seen several ways. For many years, it common to use the default *he* pronoun when indicating a hypothetical person, such as in Example 38:

[38] When a person calls, and you are taking a message, ask him for his name. (Constructed)

While the use of the default *he* pronoun has recent fallen out of favor, it is a clear example of how *male* is viewed as the default sex within the pair.

*Female* also implies a purely biological description of an entity, similarly as would be applied to an animal. It lacks the semantic property +HUMAN that *woman* has, which increases the potential to have a pejorative meaning. Additionally, the word *female* is often used as an adjective to mark what is perceived as an odd variation from the norm, especially related to

occupation<sup>9</sup>. Consider the following constructed examples:

[39] Female athletes work hard.

[40] #Male athletes work hard.

[41] Athletes work hard.

As athletes are commonly assumed to be men, speakers often use *female* to modify the noun phrase when talking about athletes who are female, as in Example 39. However, if a speaker is talking about a group of male athletes, such as the Dallas Cowboys, Example 41 would likely be the sentence chosen rather than 40. Example 40 would seem semantically strange or marked unless the conversation is specifically comparing male athletes to female athletes. Because *female* functions as the marked sex, it is also on the negative end of the spectrum and *male* is at the positive end.

Lastly, the negative bias in *illegal* is clear within the meaning, that is something that is against the law. Therefore, when *illegal* is used as an adjective, it modifies the noun implying behavior associated with criminality. However, *illegal* is restricted to abstract nouns, actions, or objects and rarely human beings. Consider the following examples:

[42] The driver made an illegal left turn.

[43] He was caught with illegal drugs.

[44] #Bob was arrested because he is an illegal businessman.  
(Constructed)

Although the adjectival use of *illegal* in Examples 42 and 43 work perfectly, Example 44 comes across as semantically strange. While someone might be able to force a meaning of *an illegal businessman*, it is not the way a speaker would naturally describe the referent. Actions

---

<sup>9</sup> A similar pattern can be observed with *male* for certain occupations that are considered dominantly female. For example, *male nurse*.

and objects are illegal, but people are not.

Now contrasting this with *legal*, which implies something that follows the law, it becomes apparent that *illegal* is heavily on the negative end of the scale while *legal* would be on the positive end.

Considering the negative meaning in the adjectives in question, the negativity is exponentially amplified when the adjective is used with nominal function. This dehumanizes the referent by removing the noun and thingifying the adjective in its place. Now the referent is simply a single property or quality that has a base negative meaning. This will be further discussed in Section 3.4.

### 3.3 Contrasting in Pairs

Each of the forms in question has another term serving as its contrasting pair: *illegal/legal*, *poor/rich*, *gay/straight*, and *female/male*. Both sides of the contrasting pairs can nominalize and are found as nominals in my corpus. However, this does not mean they are equally pejorative when nominalized. As adjectives, the contrasting pairs function on a positive/negative scale in relation to its pair. The adjectives in question *illegal*, *poor*, *gay*, and *female*, are on the negative end of the spectrum while their contrastive pairs, *legal*, *rich*, *straight*, and *male* are more on the positive end, as was discussed in Section 3.2. Additionally, how the nominalized contrastive pairs are used together reveals an important distinction between *illegal*, *poor*, *gay*, *female* on the one hand and *legal*, *rich*, *straight*, *male* on the other.

While the nominalized constructions of *rich*, *legal*, *male*, and *straight* might come across as strange (marked) and therefore possibly perceived as having a negative meaning, these examples have a less strongly pejorative meaning than the nominalized *poor*, *illegal*, *female*, and



*gay*. Recall that in the introduction (1.1) I discussed that derogatory language functions by focusing on one specific characteristic in a negative way. Similarly, each of the adjectives in question have a socially stigmatized characteristic that becomes the focus due to nominalization. On the other hand, with the contrastive pairs *rich*, *legal*, *male*, and *straight*, the adjectival characteristic that is nominalized is not stigmatized.

In comparing the use of nominalized *rich* to *poor*, the characteristic highlighted in both has to do with wealth and class. While one term, *poor*, highlights a stigmatized element of the group, the other term, *rich*, lacks the same stigmatization. In fact, it highlights a desirable and advantageous characteristic.

When non-pejorative forms such as the nominalized *rich*, *legal*, *male*, and *straight* appear in the corpus, the function of these forms appear to be mostly as a means of contrast to their counterparts. *Legal* is a prime example of this. While several examples of this form exist in the corpus, it has been within the context of a conversation about *illegals* and not within a conversation solely about legal immigrants. The point being that while these nonpejorative constructions exist within the corpus, the marked nominal form seems to be mostly for parallelism of form as seen in Example 45:

- [45] **S1.** legals hate illegals.  
**S2.** Most legals don't approve of illegal immigration.  
**S3.** Sadly this isn't always the case, because a lot of those legals used to be illegals.  
**S4.** Some of those illegals give birth to many many legals, and indoctrinate them.  
(Reddit, 2016)

In the above examples, the nominal *legal* is used as a contrast against the actual topic at hand, *illegals*. In making statements that *legals* hate or don't approve of *illegals* (S1, S2), the term seems to function in reinforcing the undesirable nature of being *an illegal*. The other contrast

seen here, with S3 and S4, is that *legal* might be disparaged themselves but only by being tied back to the origin of illegal status.

Likewise, *rich*, *male*, and *straight*, are used as comparisons with the target forms *poor*, *female*, and *gay*, as seen in Examples 46-48 below:

[46] If a poor does its “Cheap” !! If a rich does its a “Trend”  
:/ (Twitter, Sept. 26, 2014)

[47] Until the government has codified “separate but equal”  
facilities for gays and straights then there has been no  
oppression. (blacknright.wordpress.com, 2009)

[48] the gays are so selfish like the straights just want the  
attention their used to and they don't even get a pride flag  
(Twitter, Oct. 3rd, 2016)

Even with this contrasting nonpejorative nominal form, it was sometimes a challenge to find nominal uses of the non-stigmatized adjectives, as often comparisons were made in the adjectival form such as in Example 49:

[49] I've spoken to a gay and he says only a very very few gays  
like being girly but he said straight guys are the one who  
are really girly but then don't show it out cause they like  
it that way. (Twitter, Aug. 13, 2016)

In Example 49, *gay* is used as a nominal twice, but the speaker does not parallel the nominal form when using *straight*. This reflects a more common tendency for the contrastive pair to remain an adjective while *illegal*, *gay*, *female*, and *poor* are nominalized.

Furthermore, the pejorative use of *female* is more complex as it has two contrastive pairings beyond the expected pair *male*. Often, this pejorative use of *female* is instead paired with *man*. Consider Examples 50 and 51:

[50] A man needs a female like a fish needs a lobotomy (YouTube  
commentary, 2016)

[51] when a man is thinking about getting into a marriage or just a long-term relationship with a female we're gonna grade you on a scale of 1 to 10, okay? (Elite, 2010)

This uneven contrast of *female* and *man* seems to heighten the pejorative meaning of *female*, as *female* does not have the contrasting semantic properties of *man*. *Man* has the semantic properties of +HUMAN, +ADULT, +MALE. Likewise, *woman* has the properties +HUMAN, +ADULT, +FEMALE. However, *female* only contains the property +FEMALE BIOLOGICAL SEX. There is no property of +HUMAN in *female*, as *female* can be used with all animals. Therefore, contrasting *female* with a noun that has this +HUMAN quality only amplifies the lack of this property in *female* and thus amplifies the pejorative meaning.

Along with *female* being contrasted with *man*, *female* also has a second contrasting pair with *woman*. While *female* and *woman* seem as if they should be semantically similar, these contrasts draw distinct boundaries as to which individuals qualify as *women* and which individuals qualify as *females*. Consider Example 52:

[52] Money attracts the female you want, struggle attracts the woman you need. (anonymous quote, various sources)

In 52, *female* and *woman* are two distinct groups that do not overlap. The distinction made here is based on behavior, so if someone is greedy, they would fall into the *female* group but if someone is loyal they would fall into the *woman* group. Again, these contrasts only amplify the property differences between *female* and *woman* which heightens the pejorative meaning. These two distinct pairing contrasts of *female* will be discussed in further detail in Chapter 4.

### 3.4 Nominalization

#### 3.4.1 The Zero Plural, Nominalization, and Adjectives with 'Empty' Nouns

Throughout this thesis, the adjectives in question are often referred to as undergoing nominalization. In this study, nominalization refers to adjectives being used as nouns and being able to pass syntactic tests for nouns such as being able to take a plural marker, be possessive, and take determiners.

In Section 3.1, the zero plural form (*the poor are*) is introduced as being the closest thing to a nominal form available for some adjectives. In fact, it has been shown that the zero plural behaves more like an adjective than like a noun., Likewise, the indefinite and definite singular and the bare and definite plural forms of *illegal*, *gay*, *female*, and *poor* behave more like nouns.

Günther (2016) proposes that zero plural form is not a noun but an ellipsis of adjectival form, or rather an adjective with no noun. Günther argues that zero plural forms, such as *the poor*, are actually adjectives with what she calls a silent empty noun. Typically, in English an ellipsis uses an anaphoric *one* such as *the rich ones*, but in the silent empty noun forms, *the rich*, the anaphor does not surface. Günther also notes that the zero plural forms behave more like adjectives than nouns morphologically and they do not take on plural markers, but they can take on comparative and superlative morphemes such as *the wealthier* or *the poorest*.

Ross (1967) and Pullum (1975) also propose that these zero plural forms have a head noun underlyingly, though they differ on what that noun is. Ross argues that it is an anaphoric *one*, such as *the strong (ones)*, while Pullum states that the anaphoric *one* is not restrictive enough as it would allow a larger number of possible referents than the more limited *the strong*. If the underlying form is *the strong ones*, this would mean that the referent could be human

beings as well as objects such as bridges, rubber bands, and nylon ropes as seen in Example 53.

However, Pullum argues that *the strong* is restricted to human reference as in Example 54.

[53] Look through this box of rubber bands and pick out the strong ones. (Pullum, pg. 175)

[54] It is the strong that survive the Thunderdome. (Constructed)

Additionally, Pullum argues that the anaphoric *one* does not work for all adjectives. Zero plural forms like *the known*, *the inevitable*, and *the supernatural* have an abstract, nonhuman meaning which makes the use of *ones* in the underlying form ill fitted.

While Pullum doesn't propose a head noun that would work for all zero plural forms, he does propose a rule for this process, which he calls the *people deletion rule*. If the referent of the zero plural form is human, *people* can be used as the head noun in the underlying form. The surface level form is explained simply by the deletion of the head noun, *people*, and the article and adjective remains.

Pullum further explains that this rule proves that the zero plural form is actually an adjective and not a noun. To show this, Pullum argues that these forms can take on adjectival affixes (55), can be modified with adverbs (56), can be modified with adjective intensifiers (57), and cannot take the nominal plural affix -s (58), cannot take possessive markers (59 and 60).

[55] The stronger protect the weakest in the community.

[56] The really strong

[57] The very strong

[58] \*I've been doing a comparative economic study of the poors of different countries.

[59] The houses of the poor aren't as interesting to visit as those of the rich.

[60] \*The poor's houses aren't as interesting to visit as the rich's. (Pullum, p. 175)

Here Pullum's discussion touches on two aspects that relate to the target forms discussed in this thesis. While the zero plural forms cannot take the plural marker (58), the innovative forms in question can (61).

[61] Poors are why everything costs so much nowadays. Gotta pay for losses due to thefts and entitlements. (Twitter)

Likewise, Example 60 above points out that the zero plural forms cannot take a possessive marker. While this is true for the zero plural forms, the sentence in 60 could work as one of the innovative forms with some mild adjustments. For example, as in 62 below, if *the poor* and *the rich* were singular instead of a zero plural and the sentence had appropriate singular agreement, the sentence would work within the scope of these innovative forms. Likewise, *the poor* and *the rich* could be changed to the definite plural form and the sentence would be grammatical as in 63.

[62] The poor's house isn't as interesting to visit as the rich's.

[63] The poors' houses aren't as interesting to visit as the richs'.

Pullum's argument is correct when restricted to the zero plural. However, the argument does not extend to these innovative count noun forms. Just as Pullum used plural and possessive markers to prove the zero plural forms were not nominalized, the same argument shows that the forms in question are in fact nominalizations. Looking at the phrase *the illegals*, Pullum's *people deletion rule* will not work because the underlying form would have to be *\*the illegals people*.

While there are indeed similarities and relations between the zero plural forms and the innovative count noun forms, the main distinction is the zero plural does not undergo nominalization while the forms in question do. The next section discusses the semantic impact of this nominalization process and how this strengthens the pejorative meaning.

### 3.4.2 Nominalization

When a word changes syntactic category, there is a slight shift in meaning. Therefore, understanding the semantic shift that occurs in typical adjectival nominalization is relevant to understanding the shift that occurs in these innovative forms. Additionally, understanding the semantic differences between adjectives and nouns furthers our understanding of why the forms in question take on a pejorative meaning.

Wierzbicka (1986) states that an adjective merely attributes one property, whereas as a noun it references a kind with various properties entailed within that kind. For example, as an adjective, *blonde* attributes only one property: the color of hair. When this adjective is converted into a noun it becomes a kind which entails various properties: *female*, *dumb*, *sexy*, as seen in Figure 3.5.

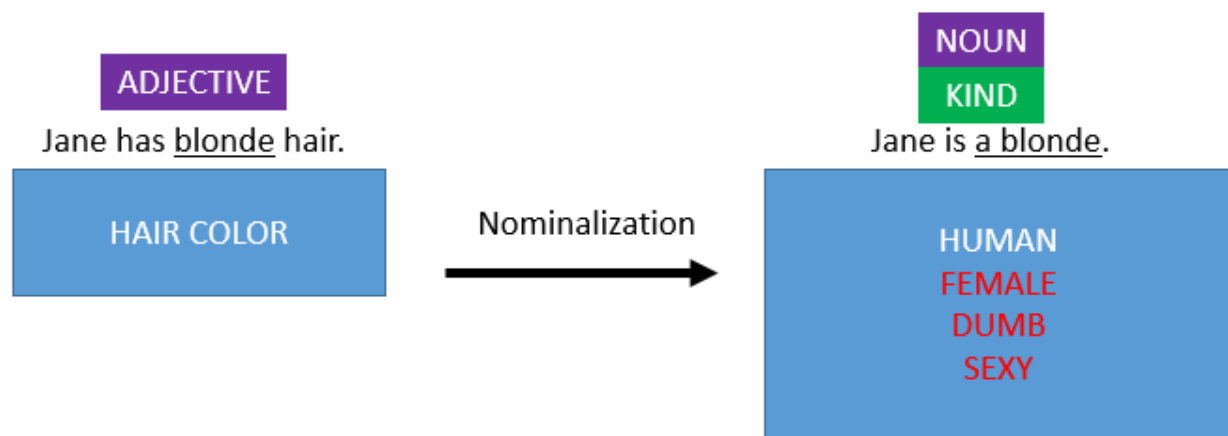


Figure 3.5: Creation of a Kind During Nominalization: *Blonde* (Wierzbicka)

Many of the entailed properties of this kind are stereotypical properties that are associated with the prototypical concept of *a blonde*. While it is perfectly possible for a man to be a blonde, it would not correlate with the prototypical concept. Wierzbicka argues that a sentence such as *She*

*is married to a blonde* comes off as slightly marked based on the implied gender of the referent, which conflicts with the gender associated with the prototypical *blonde*.

Applying Wierzbicka's insights to the current data, the semantic shift that occurs with the adjectives in question becomes more apparent. Taking *illegal* as an example, as an adjective, *illegal* simply entails that something is against the law. However, as a noun, based on a prototypical concept, *illegal* entails *foreign*, *criminal*, *freeloader*, and *Mexican* as in Figure 3.6.

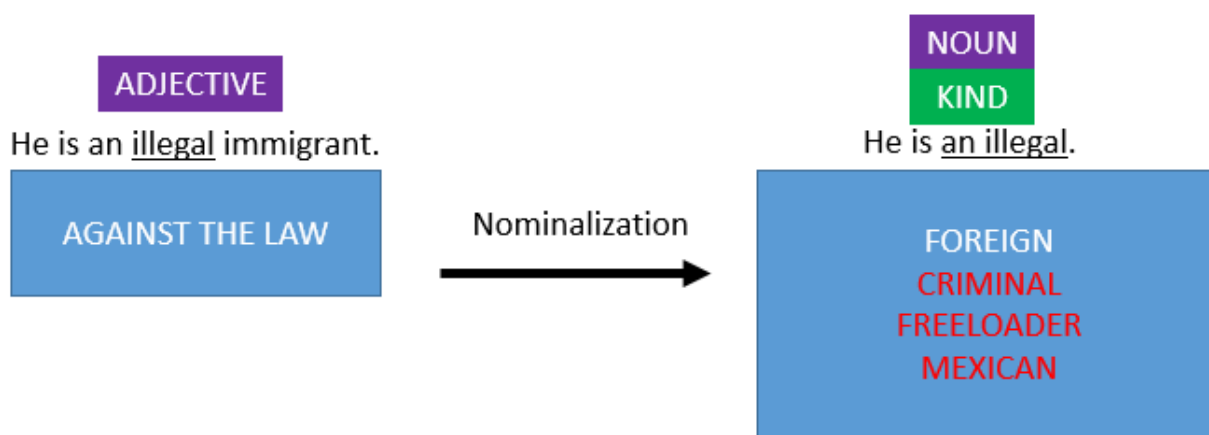


Figure 3.6: Proposed Kind Creation for *Illegal*

This is a prime example of stereotypical ideas tied to the prototype that occurs during the nominalization process. While *Mexican* might seem too specific for the basic definition of *illegals*, I argue that it is part of the prototype of *illegals*. *Illegal* can and is used to reference immigrants from other countries, but the current political discourse strongly ties the use of *illegal* to topics relating to Mexico, such as Donald Trump's plan to build a wall along the Mexican border. As seen in Example 64 below, the indefinite singular *illegal* is characterized with murderous, criminal acts, which is then cited as a reason for deportation and building the wall along the border to Mexico.

- [64] Any American citizen killed by an illegal is enough to deport all and build a wall! nothing to do with being racist (Twitter, Jul 30, 2016)



It is the negative content in uses as in Example 64 that fix the prototypical idea of *illegal* as Mexican. How the form is used, and the topics associated with the term create the prototypical idea of the term. In many ways, a legal immigrant from Mexico is referentially closer to the prototypical idea of *an illegal* than an undocumented immigrant from a country such as Norway.

Similarly, *females*, *gays*, and *poors* have prototypes that are grounded in how the terms are used and the social stigmas attached to the kind. This development of a prototype with stereotypical properties along with a generic use of these forms fits perfectly into the function of derogatory language. If an individual is labeled as a gay, it is the prototype of the kind that is brought to mind, not the individual. When these nominal forms are used to reference a specific individual, it is the kind that is invoked, and the individual is disregarded.

### 3.4.3 Adjectives and Nouns

Wierzbicka argues that, in some circumstances, there is a tendency for speakers to choose an adjective form rather than a noun form as a means of being polite and less direct. For example, the Japanese word for cripple, *izari*, is a noun. However, as *izari* attributes a stigmatized characteristic, there is a tendency to use *izari* as an adjective as a means of softening the stigma. This phenomenon of adjectives being less direct and nouns being more direct seems to occur in English as well. Using the same word, *cripple*, the tendency would be to use it as an adjective (65), unless someone wished to be more blunt or offensive.

[65] The crippled man needed help.

[66] The cripple needed help. (Constructed)

Wierzbicka also explains that when this type of adjectival nominalization happens, the nominalized form can typically reference human beings rather than animals or objects. While

*illegal* as an adjective usually modifies inanimate objects or abstract nouns, such as in Example 67, when it is nominalized it can only reference human beings as in Example 68 and not objects, as in Example 69.

[67] He took illegal action against the group

[68] His cousin is an illegal.

[69] \*John sells watches, but he only sells illegals. (Constructed)

Bolinger (1980) argues that generally adjectives are words that carry bias and often exist in biased pairs, such as old/young, tall/short, fat/thin. Likewise, Bolinger states that nouns should be less biased as they are representations of reality. A shoe is a shoe, a book is a book, there are no elements of degree. However, Bolinger points out how easily nouns can take on bias and more importantly, when they do how much more potent that bias is in noun form. Compare the following examples:

[70] Jill fusses.

[71] Jill is fussy.

[72] Jill is a fussbudget. (Bolinger, 1980, p. 79)

In Examples 70-72, note that the verb (70), adjective (71), and noun (72) forms of the same base word *fuss*, but the degree of bias shifts between each sentence. In 70, Jill fusses within the moment. In 71, *fussy* is applied as a quality to Jill and therefore takes on more bias in comparison to the momentary application in 70. However, an even higher level of bias is seen in 72 with the noun *fussbudget*. No longer is *fuss* just a quality but it is now it is applied with the branding and permanence of a noun.

Following Bolinger's argument, the permanency of nouns is a big part of what makes them far more powerful when their meanings are informed by stereotypical bias. The qualities

applied through adjectives have degrees and therefore they reflect assessments, whereas nouns are fixed. As Bolinger explains, “if we call her an *ingrate* we put a brand on her: the noun implies that the world puts people like this in a class by themselves” (p. 79). In other words, in choosing a noun form over an adjective form, the speaker is labeling the referent in a way that cannot be removed like a quality can. Likewise, the referent is categorized and put into a class which corresponds to the kind Wierzbicka discusses, and with that class come associative properties.

Bolinger explains how these brands can work as a syllogism, implying far more than just one proposition. He uses *female* as an example, calling it a derogatory term in everyday conversation. His example is as follows:

[73] **Example:** Did you see that female try to cross the street ahead of me?

**Syllogism:** Did you see that person...? Said person is a female. Females are (stupid, unreliable, troublesome, etc). Therefore, said person is stupid (unreliable, troublesome, etc.) (p. 78)

Example 73 is the actual statement made and what follows is the possible syllogism Bolinger suggests. Note that saying *female* instead of a more neutral *person*, a whole list of possible attributes associated with that brand are now applied to this specific referent. This level of subtle bias is one possible reason why the interlocutor might choose to say *female* over *person*, much like a frustrated speaker would choose *idiot* over *driver* in “This idiot just cut me off.” Branding a referent with *idiot* expresses much more desired bias than *driver*.

### 3.5 Reference

#### 3.5.1 Specific Reference

Variation with reference also influences pejorative meaning in the terms in question.

Wierzbicka states that these types of nominalizations are more restricted in use as they do not feel fully noun-like. She explains that certain nominalizations, such as *Blacks*, are less restricted, as they feel more like a noun than other nominalized types, such as *illegals*. She argues that because of this lack of nouniness, the forms would likely be restricted to generic reference, as in Example 74, and specific reference form, as in Example 75, would be seen as ungrammatical.

[74] The illegals tend to take jobs that nobody else would accept.

[75] \*The illegal, caught by the police, started to cry.  
(Wierzbicka, 1986, p. 366)

However, in recent data, it appears these classes of nominalization are being used for both generic use and specific reference, including, as shown below, forms of *illegal* (76), *poor* (77, 78), and *gay* (79):

[76] The illegal ran and was later caught... our son had to look him in the eyes and identify him... makes you feel uneasy!  
(Twitter, April 25, 2010)

[77] LOL so a Poor wants to run for President? What does Sanders know about the common American then. (Twitter, April 15, 2016, A poor referencing Bernie Sanders)

[78] Ew, isn't that what poor use for phones or something. How awful that a poor stole your CC info. Such is the state of our country. (Twitter, Nov. 23, 2015)

[79] The gay said, 'Do we know her from the community center.' And the transwoman said 'No, bitch; we robbed her.' (Twitter, May 31, 2016)

While all of these examples do reference specific individuals, there is still an element of generalization present. In essence, the individual is labeled as the kind rather than acknowledged

as an individual. Overall, these classes of nominalization are becoming less restricted than what Wierzbicka originally stated and therefore more noun-like to the speakers who use these forms. This shift in reference can influence the negative impact of the nominalized terms and helps explain how certain grammatical forms are more pejorative than others. Consider Figure 3.7.

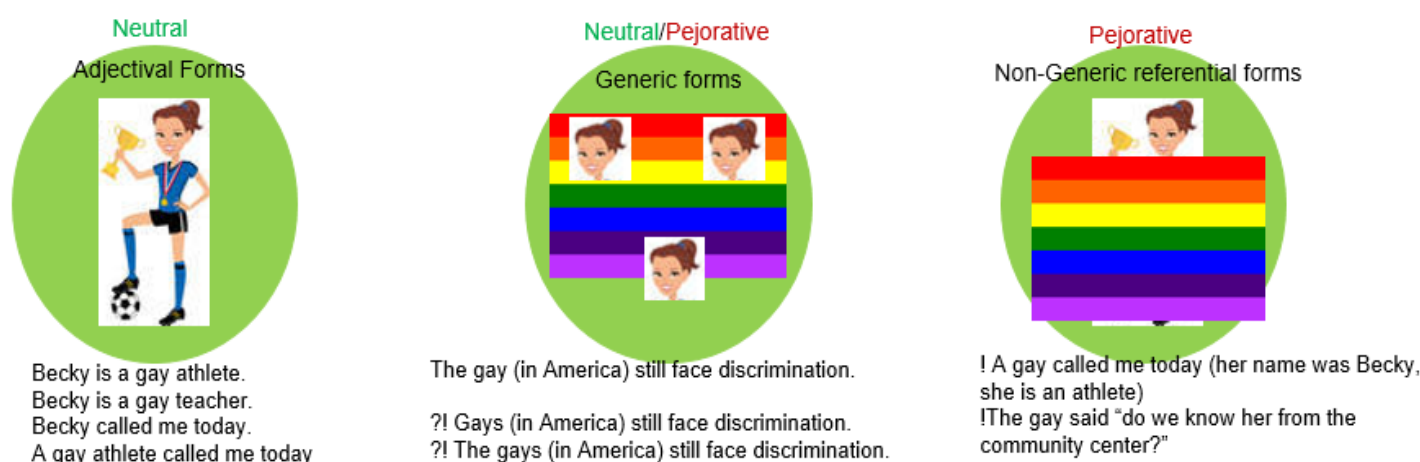


Figure 3.7: Reference Spectrum

In the first circle, the adjective form gives a neutral read. In the sentence *Becky is a gay athlete*, *gay* is only a single property attributed to Becky. She is also an athlete and more importantly she has a name. This allows the adjective term *gay* to give a neutral meaning. In the second circle, there are examples of generic reference forms, which will be discussed in 3.5.2. With genericity, there are degrees of generic meanings that vary based on the form chosen. Likewise, generic reference can be either neutral or pejorative, depending mostly on the grammatical form. The zero plural form is one of the possible forms used for generic reference. For example, the sentence *The gay (in America) still face discrimination* uses a zero plural form. This form does not produce the pejorative meaning like the forms in question. Comparing the definite plural form of this sentence *The gays (in America) still face discrimination*, the grammatical form could produce a subtle negative meaning as the definite plural can signal distancing and nonmembership, which is discussed further in Section 3.8.

The last circle contains generic/specific reference examples which has a pejorative meaning for the terms in question. In the sentence example *A gay called me today*, the fact that the individual who called is named Becky or other facts about Becky, such as she is an athlete are completely erased. As the reference term is entirely focused on the single adjectival property of *gay*, the term dehumanizes the referent which amplifies the pejorative meaning. This is made clear by the illustration, as the individual Becky can no longer be seen as the property *gay* takes the focus.

The next section further explains generic reference, specifically when a kind is referred to in a generic manner.

### 3.5.2 Genericity: Kind-Referring

In this thesis, two types of reference are discussed, generic reference, where a kind or type is referenced, and non-generic reference, where specific entities are referenced as discussed in Section 3.5.1. Whether a speaker is using the nominalized terms in question to address specific individual entities or a generic kind is important, as reference reveals the way speaker views the referent. Recall that Wierzbicka argues that the nominalized forms should be restricted to generic use because the forms do not fully feel like nouns. However, based on the corpus data, *illegal*, *poor*, *gay*, and *female* have evolved and have much more flexibility in their ability to reference specific individuals. The nominals might be marked, but speakers who use them treat them as full nouns and do not restrict them only to generic use. In other words, analyzing the referential aspect reveals part of the progression of the nominals.

With the importance of the referential type in mind, genericity must be explored for its role in the pejorative process. Genericity is an overarching term for the use of either a noun

phrase or the whole sentence to discuss topics in a generalizing manner. Carlson & Pelletier (1995) explain that kind-referring genericity is the use of noun phrases to reference a kind or category of thing rather than referencing a specific entity. This type of genericity is commonly used to make a general statement that is meant to characterize an entire kind. The following are standard examples of kind-referring genericity:

[80] Lions live in Africa.

[81] The dog is a curious creature. (Carlson & Pelletier, 1995)

In Example 80, *lions* does not reference a specific group, such as 10 or 20 lions, but rather the species of lions. Likewise, Example 81 references the canine species, even though the grammatical form is singular.

There are different grammatical forms possible when using kind-referring genericity which are discussed below. These forms correlate with the innovative grammatical forms under investigation. Linguists have noticed that certain forms were used generically in certain contexts and that there appeared to be a subtle semantic difference between them.

In examining generic meaning, Lawler (1973) categorizes the three grammatical forms possible as *prototype* (definite singular-*the cat*), *norm* (bare plural-*cats*), and *definition* (indefinite singular-*a cat*). With the *prototype* (definite singular), the conceptual idea of the kind being referenced is the prototypical image of that kind. The *norm* (bare plural) is a more general and personal concept of what characteristics might be associated with the kind. Lastly, the *definition* (indefinite singular) can only be used to reference characteristics that are absolute necessities to the kind. For example, in 82 below, the indefinite singular form is allowed, because *polyphonic* is a required property of the madrigal. In comparison, Example 83 is

deemed ungrammatical, because *popular* is not a required property of madrigal. However, popularity can be a required property of a kind, seen in Example 84.

[82] A madrigal is polyphonic.

[83] \*A madrigal is popular.

[84] A football hero is popular. (Lawler, 1973)

When defining generic reference, the literature does not typically discuss the use of the definite plural form. This is due to the specifying aspect of the definite article, which leads to an interpretation that is usually specific. Consider Example 85 where the definite article functions to specify a specific group of individuals, yet *the illegals* can also be read as kind-specific genericity.

[85] Yeah but the illegals get immediate healthcare without wait.  
WTH happend to America past 8 years???? (Twitter, Oct. 6, 2016)

But what is the specific group being referenced here? While it is possible the implications are *the illegals in America* which would be more specific, separating this reference from the innovative kind *illegal* becomes rather difficult. Consider the examples below where both bare plurals and definite plural forms are used.

[86] (a). We The Citizen American working man/woman are becoming slaves 2 the illegals! Our \$/benefits go to them. (Twitter, Oct. 5, 2016)

(b). We The Citizen American working man/woman are becoming slaves 2 illegals! Our \$/benefits go to them. (Constructed)

[87] (a). Currently on a bus like some poor, I want to make it clear that I am not a poor, I am extremely wealthy, I just take buses to view the poors. (Twitter, July 6, 2016, might have a sarcastic read)

(b). Currently on a bus like some poor, I want to make it clear that I am not a poor, I am extremely wealthy, I just take buses to view poors. (Constructed)



While the definite plurals do bring about a specific reference read, this is due solely to the definite article. The referent itself, whether it is underlyingly the illegals in America or something else, does not seem to be relatively more specific than the kind created in these innovative forms. Instead, it is often the case that these definite plurals have a dual reading of both specific and generic. This split between a kind-referring generic interpretation and a specific reference interpretation complicates the semantic meaning. In Section 3.8, a possible explanation for the prominence of definite plurals in the data is discussed, as the grammatical form of the definite plural also functions as a means to imply non-membership in a referent group.

The next section further discusses prototypicality and its role in how the use of *illegal*, *female*, *gay*, and *poor* is deeply rooted in stereotypes.

### 3.6 Prototypicality

While prototypes were briefly discussed earlier in Section 3.4, the concept is a complex one and requires a more in-depth discussion. Carlson and Pelletier (1995) explain that the prototype approach is based in cognitive psychology (Rosch, 1978; Platteau, 1980; Nunberg and Pan, 1975; Heyer, 1985, 1987, 1990). It is a useful concept when applied to generic reference as it helps explain how generic sentences framed within a universal statement could be accepted semantically, such as in Example 88.

[88] A cat has a tail. (Carlson & Pelletier, 1995, p. 46)

An argument can be made that not all cats have a tail. However, in sentence 88, it is acceptable to state generically that *a cat has a tail* using a universal statement structure. The prototype approach explains that sentence 88 does not mean (*every*) *cat has a tail* but instead that *a*

(*prototypical*) *cat has a tail*. In other words, semantically this sentence is acceptable within a universal structure because it is understood to be a reference to the prototypical image of a cat, and having a tail is part of that prototype.

Despite the convenient explanation of generic statements, Carlson and Pelletier reveal a problem with using the prototype theory as a sole explanation for this phenomenon.

[89] A duck has colorful feathers.

[90] A duck lays whitish eggs. (Carlson & Pelletier, 1995, p. 47)

Examples 89 and 90 are both semantically acceptable, just as is Example 87 above. However, only male ducks have colorful feathers and only female ducks lay whitish eggs. Carlson and Pelletier view this as problematic in terms of the prototype because it cannot be attributed to one example of an ideal duck.

Carlson and Pelletier offer up a secondary approach that is related to prototype approach, that the reason these sentences are allowable semantically is because they are expressing stereotypical aspects or prototypical features of some duck, like the cat in Example 84. As stated earlier, Wierzbicka (1986) argued that stereotypical properties are associated with a kind when it is formed, such as with the nominalized *blonde* (dumb, sexy, female). Therefore, these stereotypes appear to function as a part of the prototype itself.

As generic sentences often are structured as universal statements, the prototypical concept of a kind allows for accepted generalizing statements that can have exceptions, as long as they are part of the prototype. However, these prototypical ideas do not always correlate with what is most likely true. Consider the following sentences from Carlson & Pelletier (1995):

[91] A lion has a mane.

[92] #A lion is male.

[93] Chickens lay eggs.

[94] #Chickens are hens. (Carlson & Pelletier, 1995, p. 48)

As generic statements, sentences 91 and 93 would be preferred and sentences 92 and 94 would be perceived as incorrect. However, sentences 92 and 94 have a higher chance of being true than 91 and 93 because there is a 50% chance with gender, but characteristics like *having a mane* and *laying eggs* are not only gender restricted but age restricted as well. The reason sentences 91 and 93 are acceptable generic sentences is due to the prototypical idea of *lions* and *chickens*. Prototypically lions do have manes and chickens do lay eggs, but gender is not a part of the prototype. This concept of less true statements being allowable within the prototype is strikingly similar to how stereotypes work. Generalizing statements can stray from the truth as long as they are part of the prototypical concept of the kind.

[95] Why don't gays like being girly? Cause a gay is normally called girly. (Twitter, July 25, 2016)

[96] but cat food is and that's all the poors deserve to eat! (Wonkette.com commentary, April 27, 2015)

[97] Money attracts the female you want, struggle attracts the woman you need. (anonymous quote, various sources)

[98] A lot of illegals pop out kids like crazy too, Youll see them with 5 toddlers all of the time when you take public transport. (Reddit, 2015)

As seen in the examples above, genericity and prototypical concepts of a kind easily tie in with the use of stereotypes. In Example 95, the use of *girly* alongside *gays* brings to mind a prototypical image that gays are girly. In Example 96, *poors* the genericity is used to loosely compare the poor to animals. Example 97 attributes greed to *females* and Example 98 stereotypically claims that *illegals* have too many children. The contexts in which a term is used lead to the development of the prototypical idea.

The next section further discusses the importance of grammatical form but specifically highlights how the definite plural can be used to imply negative social meanings toward various groups.

### 3.7 Markedness

I have discussed the more standard grammatical form of the zero plural, such as *the poor are*, in Section 3.4.1. Additionally, I have shown innovative forms that can occur in adjectival nominalization, such as the indefinite and definite singulars, and the bare and definite plurals as discussed in 3.1. With each innovative grammatical form possible, there appears to be a degree of innovation in terms of how similar or dissimilar the form is from the zero plural. Innovation in grammatical forms correlates with markedness in that the more innovative the form, the more marked it is. Likewise, markedness in these cases appear to correlate with pejorative meaning in that the more marked a construction is the more pejorative it is. With variation in language, markedness allows for a means to assess the relationship between standard and innovative forms.

As each of the innovative nominalized forms seem to be slightly different semantically, a markedness hierarchy might play a role in the semantic variation. Croft (1991) explains: “Formally, one can represent the markedness pattern of paradigmatic elements of a category as privileged members of a single set, the set of members of a grammatical category” (p. 54). To put it another way, markedness is not binary, instead it is a hierarchy with varying degrees of markedness relative to other elements within the category. To demonstrate his point, Croft explains that there is a markedness with plurality. When comparing dual, plural, and singular, plural is marked relative to singular form; however, plural behaves as unmarked relative to dual. This would be represented as such: singular > plural > dual.

Similarly, the variation of grammatical forms within the innovative nominalizations are marked along a spectrum relative to each other. Taking into consideration the proposed process, the following markedness hierarchy is proposed: adjective > zero plural > bare plural > indefinite and definite singular > definite plural. The bare plural appears to be the least marked of the innovative grammatical forms and likewise likely to be perceived as the least pejorative. This is due to the bare plural being restricted to generic use and therefore closer to the use of the zero plural. Overall, the bare plural has less referential flexibility than the indefinite and definite singular and the definite plural. As markedness seems to relate to pejoration, this hierarchy is useful in hypothesizing the pejorative differences between forms.

While there does seem to be a connection between grammatical form and degrees of pejorative meaning, this thesis does not attempt to pinpoint which grammatical form is perceived as more pejorative, instead it focuses on the reasons why this pejoration occurs in connection to the grammatical shift. However, based on observation of the data, I propose that it is likely the level of pejoration might follow the proposed markedness hierarchy. Analyzing the levels of pejoration between innovative grammatical forms would be an interesting future study to see if the markedness hierarchy does indeed correlate the degrees of pejorative meaning.

### 3.8 Grammatical Form and Associative Content

As mentioned in Section 3.7, the definite plural form appears to be the most marked and likely the most pejorative as well. According to Acton (2014), this is due to an otherness that is pragmatically entailed in the grammatical form. Acton explains that the bare plural form, such as *Americans*, functions open-endedly, meaning it does not imply the interlocutor's membership or non-membership to the group mentioned. However, using the definite article form, *the*

*Americans*, specifically signals the interlocutor's non-membership to the group. Not only is it just a matter of the interlocutor not belonging to a group, but the grammatical form itself seems to function as a way for the interlocutor to bluntly make it known that they are not a part of the group.

[99] The illegals will break your piggy bank. Then come for more.  
(Twitter, Sept. 26, 2016)

[100] Good morning can't wait to see what the gays are gonna be  
mad about today (Twitter, Sept. 29, 2016)

As seen in the examples above, both uses of the definite plural form serve to signal that the speaker does not belong to the group mentioned. Additionally, the definite plural grammatical form entails distancing. This is the clearest when the nonmembership of the interlocutor is already known. Consider Acton's example below:

[101] [Canadian mother to Canadian daughter, concerning her  
daughter's husband—a U.S. citizen who is eating a Big Mac at  
9am]

**Mother:** He's eating a Big Mac for breakfast?

**Daughter:** What can I say? The Americans love fast food.  
(Acton, 2014, pg. 98)

In Example 101, Acton points out that because it is already known that the daughter is not American, the use of the definite plural *the Americans* over the bare plural *Americans* is not being used to signal nonmembership. Instead, it is being used to signal a social meaning and inferences by stressing her nonmembership and thus distancing herself from the reference group.

Additionally, Acton argues that there's a difference between definite plural instances such as *the Americans* and *the gays*. While there might be some negativity associated with *the Americans*, it is not necessarily perceived as pejorative. However, *the gays* seems to have an added pejorative meaning, along with the non-membership and the distancing. Acton explains that the non-membership and distancing are entailed in the grammatical form, whereas the

possible pejorative meaning comes from associative content, which is a social meaning relating to stereotypical information and associative use of the basic form.

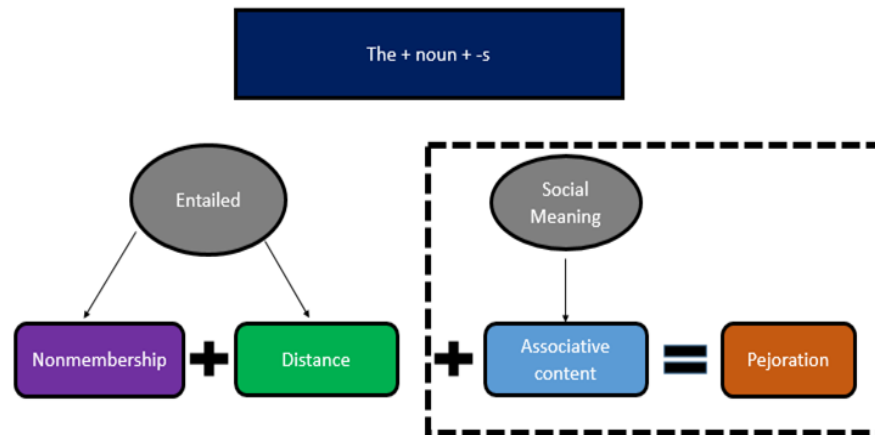


Figure 3.8: Acton's Definite Plural Nonmembership

The model in Figure 3.8 is an adaptation of Acton's definite plural nonmembership theory. As the model demonstrates, the definite plural form has two parts, what is entailed in the grammatical form and what is implied through social meaning associated with the word.

Acton's explanation correlates with the definite plural usage in my data as well.

Returning to an earlier point, *illegal* and *poor* seem to favor the definite plural forms. Taking into consideration Acton's argument, this might be due to a socially based need to imply non-membership and distance from these two stigmatized groups. That is not to say that *female* and *gay* do not experience the same distancing and non-membership. In fact, a well-known example is from Donald Trump who seems to favor using the definite plural form to reference certain groups, as seen in Example 102:

[102] For the gays out there—ask the gays and ask the people—ask the gays what they think and what they do (Donald Trump, June 15, 2016)

This example illustrates the differences between what is entailed in the grammatical form and when there is an added pejorative social meaning. In *the people*, non-membership and

distancing is entailed, but there is no pejoration. If that is compared to *the gays*, it is apparent there is something additional to the entailed meaning of the definite plural, which is the associative content. The social associative content of *gay* provides a negative interpretation of the non-membership and distancing entailed in the grammatical form. While not all definite plural forms have this pejorative meaning, it appears that the entailed meaning intensifies the pejoration associated with the term.

Acton's concept of associative meaning might also relate to why these innovative classes of nominalization take on a pejorative meaning. All four of the classes of nominalization reference minority or stigmatized groups, which tie into the concept of associative meaning. If the referent groups are often associated with negative terms and concepts, this association might be a key part of the pejorative meaning. This social meaning, along with the markedness of the innovative grammatical forms, make the uses more likely to pejorate.

It is the entailed non-membership and distance that provides the marked environment that amplifies the possible negative social meaning, such as in *the gays*. This appears to correlate with what happens with adjectives that go through the nominalization process. While the nominalization, genericity, and marked construction all help create a grammatical form that is susceptible to pejoration, the end results might produce nonpejorative terms like *a rich* or *a legal*.

Grammatical form can signal much more than definiteness and number, especially in marked forms that are prone to pejoration. When thinking about pejorative meaning, the context in which the terms are used is often important. The next section discusses the impact of context and grammatical form in the pejorative meaning of the terms in question.



### 3.9 Conclusion

In this chapter, I presented the cluster of linguistic properties that contribute to the pejorative mean of the nominal *illegal*, *gay*, *poor*, and *female*. The following is a summary of the main points covered in this chapter.

- When speakers fill in holes in the paradigms with innovative adjective nominals, the terms become marked and susceptible to pejoration.
- Additional linguistic properties contribute to the pejorative meaning including: adjective meaning, contrasting in pairs, and negative amplification through nominalization.
- Adjectival nominalization forms a kind grounded in the adjectival property.
- The kind that forms during nominalization takes on stereotypical properties associated with societies prototypical idea of that kind.
- Markedness is not binary and can vary over numerous grammatical forms.
- Both grammatical form and variation in reference type play an important part in the pejorative meaning.

In Chapter 4, I build on the theory discussed in Chapter 3 by looking closely at how the forms function in context.

## CHAPTER 4

### A CLOSER LOOK AT THE DATA

#### 4.1 Introduction

While the nominalization process discussed in this thesis can be applied to numerous adjectives, this study focused specifically on *female*, *illegal*, *poor*, and *gay*. There are various reasons why these four adjectives were chosen, but a primary factor was current events during the early stages of the study, such as the 2016 presidential election. As Hillary Clinton was a candidate for President, topics related to whether women should be President were prevalent in the media and online. Within this type of discourse and discourses directly related to Hillary Clinton, the use of *female* as a noun seemed to be much more prevalent. Additionally, with Donald Trump's platform topic of building a wall along the border of Mexico, the topic of immigration was a dominant one. Much like *female*, when investigating these topics, the use of *illegal* continued to surface. Similarly, *poor* and *gay* were found relating to political topics in general, such as welfare for *poor* and gay rights for *gay*.

Chapter 3 discusses the similarities between *female*, *illegal*, *poor*, and *gay* as they undergo the pejorative nominalization process and the reasons behind the negative meaning. In this chapter, I present some of the differences and unique features to each adjective nominal.

#### 4.2 The Pejorative Noun *Female*

Out of the four adjective nominals, *female* appeared to be the most complex. First, *female* is a noun as well as an adjective. As discussed in Section 3.1, unlike *illegal*, *poor*, and *gay*, *female* does not present any holes in the definite and indefinite paradigm. Additionally, there are valid reasons for using *female* as a noun to reference human beings, such as to include a

larger age group that includes children and adults or if the topic is gender focused. However, these uses are distinctly different from the pejorative use discussed in this thesis. The pejorative use occurs in everyday conversation when the expected noun *woman* is replaced by *female*. More specifically, the pejorative use of *female* occurs when there are no other valid motives, such as age inclusion or topic relevance, to use *female* other than to apply a pejorative label on the referent. In this section, I closely examine how this pejorative use of *female* is used and why it is distinctly different from the other uses mentioned above.

One of the most distinct characteristics of the pejorative *female* is that it occurs in contrastive pairings. Figure 4.1 indicates the expected contrastive pairings for these gendered nouns: *female/male* and *woman/man*. These pairings are indicated with a solid line below:

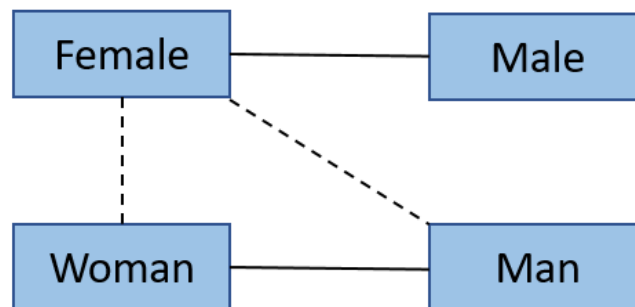


Figure 4.1: Contrastive Pairings of Pejorative *Female*

However, there are other possible contrasting pairs, *female/man* and *woman/male*. While *female/man* was a quite common contrastive pair in the data, *woman/male* was not. Lastly, the vertical dotted line indicates a relationship between two terms that are semantically similar rather than terms that can be contrasted. However, in the data collected, it appears that a contrastive pairing is being forged along the vertical axis between *female/woman*, but not between *male/man*<sup>10</sup>.

<sup>10</sup> There is potential for a pejorative *man/male* contrastive pairing, though so far not many examples have been found. *Male* could be used pejoratively similar to how *female* is being used, however, it would likely have to be

The pejorative contrastive pairings of *female/man* and *female/woman* were readily found in the corpus while a similar pejorative *woman/male* and *man/male* were not. In principle, a pejorative contrast between *woman/male* and *man/male* could exist as well, but it was not one observed in this study.

With further analysis, it became apparent that *female* and *woman* were not synonymous, and speakers used the two terms to reference two distinct groups. In a CNN interview a woman was asked about her beliefs that the president should not be a woman. In this interview, she switched back and forth between her use of *woman* and *female*. When referencing herself and the positive aspects of her life, such as being CEO, she would use *woman*. However, when she spoke of what she saw as negative behavior, such as a woman being president, she would use *female*, as seen below.

[103] I am a strong woman, I run my own company like you said but that is not the same as running the best country in the world and being commander in chief, head of state, the president of the United States, to me, should be a man not a female. (CNN Interview, Apr 22, 2015)

As seen in Example 103, *woman* is modified with a positive adjective and is used to reference the interlocutor. However, instead of just staying with the term *woman*, the speaker switches to *female* when discussing what is viewed to be objectionable behavior. Note also, *female* is used contrastively with *man* whose opposite is *woman*. Though this is one example, there are numerous examples in the data of a clear distinction between who is a *woman* and who is a *female* based on behavior. If the referent exhibits negative behavior or qualities or lacks positive behavior or qualities, they are classified as *female* rather than *woman*.

---

contrasted with *man* for the pejorative meaning to work. Example: That's not a man... that is *a male*. A beta male. A spineless liberal dweeb living in his mom's basement.

Example 104 below, is yet another example of this distinction. The speaker uses various gendered nouns, including *girls*, *women*, and *lady*, but explains that some referents do not meet the requirements for any of those gendered terms, only *female*.

[104] SOME females are just that though. They're too old and not considered girls anymore but don't conduct themselves as women and they're definitely not a lady, so they're JUST genetically females. (buzzfeed.com commentary, 2014)

In Example 104, the speaker highlights the civilized properties that are attached to terms like *women* and *lady* that *female* does not have. In order for a referent meet the requirements of *woman* or *lady*, there is an expected behavior. If the referent is instead called *female* the implications are that the referent does not fulfill the necessary qualities to be called *woman*. A semantic description of the features associated with *woman* include +ADULT, +HUMAN, +FEMALE, whereas *female* only has the property +BIOLOGICAL SEX. Additionally, the speaker presents the terms *girl*, *woman*, and *lady* on a scale that implies growth toward refined maturity. To contrast that, note the adverb *genetically* modifying *female* which incidentally also speaks to the biological reproductive function of the referent, which in this case excludes any social characteristics such as acquired social refinement. *Genetically* is also, in turn, modified with *just*. This reinforces the quality of *biological sex*. Additionally, as the terms *girl*, *woman*, and *lady* are presented on a scalar relation in the achievement of refinement, it also removes any possibility of assigning the referent a social identity.

Additionally, in contrastive pairs, *female* is often paired with *man* rather than *male*. As seen above, *woman* and *female* are not synonyms because of their semantic properties are not entirely the same. *Female* and *male* entails biological sex, whereas *woman* and *man* entail biological sex as well as socio-cultural features associated with human behavior. Therefore,

below in Example 105, it is notable that *female* is paired with *man*, instead of the expected pairing of *male*.

[105] the president of the United States, to me, should be a man not a female. (CNN Interview, Apr 21, 2015)

The contrastive use of *man/female* in Example 105 clearly implies that the word *female* entails something slightly different from the standard use of *female* and therefore no longer can be paired with *male* in the same way. Additionally, the speaker uses the *female/man* contrastive pairs to further highlight the inequality between the groups similar to the way *female/woman* are contrasted as in Example 103.

The fact that the speaker, in Example 103, uses the contrastive pairing within the topic of whether a woman should be president only emphasizes a possible motivation for presenting the two groups, *female/man*, in an unequal manner. From the speaker's point of view, one group has the necessary qualities to be president (*man*), and one does not (*female*). Therefore, if the speaker is trying to imply that the referent is unfit for office, using the term *female* contrasted with *man* emphasizes this lack of necessary qualities.

The examples discussed so far show that the meaning of *female* is acquiring a 'layer' of pejorative meaning, moving it further away from its semantic similarity to *woman*. At the same time, the data shows that *female* is becoming more semantically similar to *bitch*. Throughout the data, *female* is interchanged with *bitch*, rather than with *woman* as seen below in Example 106:

[106] Real women don't like being compared to the basic bitches you were with before, if they hurt you, they hurt you, understand not all women are going to do that to you. But I digress, back to you weak ass females, you had a good man but you treated him like shit. (YouTube.com, March 4, 2015)

In Example 106, there are two separate groups of women being referenced here: there is the group that is referred to as *real women* and then there is the group that is referred to as *basic*

*bitches* and *weak ass females*. The speaker makes a clear distinction between these two groups. Furthermore, *basic ass bitches* has the same referent as *weak ass females*, treating *females* as interchangeable with *bitches*. Additionally, equating *female* with a negative term like *bitch*, in turn triggers the use of equally pejoratively strong adjectives, such as *basic* and *weak*, and the negative intensifier *ass*.

*Bitch* is often used within pejorative slang phrases such as *basic bitch*, *thirsty bitch*, and *ratchet bitch*. While inspecting these slang phrases, it was discovered that sometimes *female* was used instead of *bitch*, as seen in Example 107 and 108:

[107] as long as your bf isn't acknowledging that thirsty female sit back and watch the dog beg for a bone ladies. (Twitter, Nov. 8, 2015)

[108] Ratchet females with brass knuckles tattoos are filth and need the shit beat out of them on a regular basis. (Twitter, Oct. 7, 2016)

When *basic*, *thirsty*, and *ratchet* are prenominal modifiers, they seem to require a negative head noun, such as *bitch* or *female*. For example, while *basic bitch* and *basic female* are used by speakers, it is not very likely a speaker will use *basic woman*. While *basic woman* is grammatically a possible phrase in English, it is not a combination that would necessarily trigger negative features because the head noun does not carry the pejorative meaning associated with *bitch* or the pejorative layers associated with *female*.

Example 109 below is an instance of the negative use of *basic ass female*:

[109] a lot of the times ladies you're not getting into long term relationships and you're not getting married because a lot of you ladies are just basic ass females. (Elite, 2010)

Example 109 is from a video blog addressing Black women about how men rate women. The interlocutor rates women on a ten-point scale with *basic ass female* being at the low end of

the scale. The speaker explains that the scale is designed to show how men rate women based on behavior and how men decide what category a woman falls into. Points are assigned or lost based on negative or positive behavior. Each point lost drew the women closer to being categorized as *basic ass females*, based on undesirable behavior as viewed by the interlocutor. Example 109 is just one instance of the interlocutor's use of *basic ass females*, a phrase that is actually used, throughout the video. Note that while he uses *ladies* in this example, he uses it to address his audience. However, the speaker is using *ladies* and *females* in contrast to reference two separate groups of women. In other words, not all members of the set of *ladies* are *basic ass females*, so the term *ladies* and *basic ass females* are not interchangeable here.

Lastly, as has been mentioned in the discussion above, the use of *female* is often tied to bad behavior based on the interlocutor's perspective. While *woman* is also used in correlation with bad behavior, it is also used in neutral contexts as well as positive ones as well. This is where *woman* and *female* differ. *Female* is predominately associated with negative behavior or contexts. As Deborah Cameron (2016) has also noted in her blog "Whereas 'woman' can feature in positive as well as negative judgments, it's hard to think of any context in which the noun 'female' is used to praise its referent: no one would say, for instance, 'my late grandmother was an absolutely marvelous female'"

#### 4.3 The Pejorative Noun *Poor*

The majority of instances with *poor* were used in the data collected in a satirical or sarcastic way. While *poor* is the least frequent of the four, (only found in written discourse), it also seems to reveal an important aspect about how well speakers understanding of how nominalizing adjectives can amplify a negative meaning. Consider Examples 110 and 111:



[110] As a proud member of the middle class, I can't wait for all of the poors to die off. (Wonkette.com, Apr. 27, 2015)

[111] As a proud member of the middle class, I can't wait for all of the poor to die off. (Constructed)

Both Example 110 and 111 have highly negative context. However, the innovative grammatical form *poor* in 110 adds something further to the instance that signals the audience should interpret it as sarcasm. However, looking at Example 111, the sarcastic meaning would not be conveyed due to the more standard zero plural form of *poor*. Additional examples below will explain this sarcastic interpretation further.

While it has been argued that *poor* is used as a pejorative satire/sarcasm tool, it is worth looking at the fine details of how it is being used. Most often it is used as a way to mimic a perceived attitude of an individual, a group, or society as a whole. In Example 112 below, the interlocutor is providing commentary on Maine Governor Paul LePage's efforts to restrict what food stamps can be used for.

[112] Gee willikers golly Jeebus on a cracker, the states of this union are stepping up their game in the eternal contest to see who can fuck the poors with the least amount of lube. (Wonkette.com, Apr. 27, 2015)

The overall tone expresses what the interlocutor believes LePage's actions are doing to the poor. While the phrase *getting fucked over* is understood metaphorically as a negative expression, the interlocutor actually intensifies this expression by tying negatively it back to a literal meaning of *getting fucked* adding *with the least amount of lube*.

#### 4.4 The Pejorative Noun *Gay*

In my analysis of *gay*, I discuss forms not included in the standard paradigm, such as *a gay* and *the gay is*. The pejorative use of *gay* is a complex form to analyze due to *gay* having an

inclusive and exclusive use, depending on whether the interlocutor is a part the gay community or is an outsider. This in itself is fascinating as the implications of the term usage seem to shift based on who is using it. Example 113 is an example of an inclusive use, as the speaker is a part of the gay community (note the use of *us*). In this example, the speaker objects to nominal use of *gay* by those outside the gay community and specifically singles out Donald Trump.

[113] Only the gays or allies of the gays get to call us “gays”, @realDonaldTrump. “Gay people would be respectful term from your type #debate (Twitter, Oct. 19, 2016)

Note that the speaker does not appear to object to the context in which *gay* was used, just that his *type* should use *gay* in the adjectival form to be respectful. Although the interlocutor’s meaning of Trump’s *type* is not fully clear, the implication is that he is not an ally of the gay community.

Reactions to this nominal use of *gay* also help expose what exactly causes this negative reaction to the adjective *gay* being used as a noun. In Example 114 below, S1 retweets (RT) someone else’s post and then S1 posts their reaction to the nominal use within the comment.

[114] **RT:** I’m so afraid of a gay

**S1:** Not that I was apart of this convo but I just LOVE when people say “a gay” like we’re some “thing.” It makes me feel so good inside. (Twitter, April 1, 2017)

S1, a gay person, touches on the odd construction of the indefinite singular use of *gay*. For them, it brings about an object or *thing* like quality rather than referencing them as human beings. This dehumanization seems to be a common thread for all four nominalized adjectives. This dehumanization relates to the lack of a humanizing noun, such as *man*, *citizen*, *person*, and instead the noun is rooted in one single adjectival property of the individual.

Looking at pejorative uses of *gay*, common themes are repeated that often relate to negative discourse on homosexuality as a whole. These pejorative uses of *gay* are often found in

exchanges, be it arguments or complaints against homosexuality often citing on biblical, psychological, or behavioral justifications. Consider examples 115 and 116:

[115] Whats worse is how much media attention the gays receive, had they not then they would never garner as much support as they have and the support they have is from people whom believe their stupid lies of “born this way” “unchangeable” (Reddit, 2013)

[116] You have no proof, and no reasons backing the “gay is natural” argument because gays are not natural and are a mental disorder which needs treatment (Reddit, 2013)

In Examples 115 and 116, variations on a common argument are brought up. The idea that being gay is a choice is often tied to arguments against homosexuality, as it means the “choice” can simply be reversed. Example 116 goes even further with this argument, tying homosexuality both to mental disorders and unnatural behavior. Similarly, Examples 117 and 118 use *gay* within the context of unnatural and deviant behavior:

[117] homosexuality disgusts me. I am not phobic by any means. I just believe it is gross. Like overstepping the lines of biology gross. I used to live 20 min from San Francisco. I can tell you that every gay that I know is a pervert. (Commentary on blacknright.wordpress.com, 2009)

[118] speaking your mind is not bigotry, if that were the case then every gay who has posted here about societies unwillingness to accept their deviant behavior would also be considered bigots. (Commentary on blacknright.wordpress.com, 2009)

In Example 117 and 118, *gay* is tied to extremely negative assessments, close to ‘name-calling.’ In 117, *gay* is equated with perversion and something that goes against natural or biology in this case. In Example 118, the interlocutor uses the term *bigotry* to turn back onto the gay posters, silencing them by calling them bigots for not accepting the labels of deviants.

Example 119 below brings about the common religious argument against homosexuality, stating that relationships are meant to be between men and women and anything that does not fit

this standard is disgusting. Additionally, note the use of both *gays* and *homosexuals* as if using both to emphasis their point.

[119] Also gays and homosexuals are going against what God wants man to be with women Homosexuals disgust me. (Commentary on blacknright.wordpress.com, 2009)

Example 120 shows a slightly different type of use for *gay*. Example 120 has a number of examples of punctuating statements used in an anti-gay blog. These are only a small portion of what was used in the blog, but these examples are a representative of the general context in which *gay* is used in this text. The blog uses a common tactic of deflecting from one social issue and thereby belittling it by presenting various other social issues that are ‘more deserving’ of attention. This blog touches on poverty, abortion, the mentally handicap, and immigration issues to name a few. After each point made, the author punctuates the point with variations of a repeated clause *then I will worry about the gays*.

- [120] 1. When those who use welfare to pay for medication they otherwise could not afford and have to choose between their life and simple self respect are no longer stigmatized, then I will ponder the so called rights of gays.
2. When the gays experience real discrimination and don't harass people who raise legitimate objections to homosexual behavior then I will worry about the gays.
3. Until the government has codified “separate but equal” facilities for gays and straights then there has been no oppression.
4. Until then I will not worry about the gays, they have proven they can take care of themselves.  
(Blacknright.wordpress.com, 2008)

These examples bring to mind microaggression as they subtly dismiss the gay community and gay rights all together without directly saying anything negative at all. However, there are implications of negative sentiment toward the gay community. In instance 1 above, the use of *so called* modifying *rights of gays* implies the speaker does not believe gay citizens deserve rights.

This is one of the few clearer indications of belittling done by the writer. The implications here being that gays do not deserve rights or at the very least they do not deserve the rights they are demanding.

In all four instances of Example 120, the interlocutor disparages the plight of the gay community. In the first example in 117, the speaker compares *gays* to another group that deals with discrimination, such as those on welfare. Similarly, in instance 3, the segregation of Blacks and Whites in America is brought to mind at the mention of the government segregating straights and gays. The interlocutor uses these examples as a means to ‘measure’ whether the gay community has truly been oppressed. Often the plight of the gay community is compared to racism in America, which is not always accepted as a legitimate comparison, yet here the interlocutor defines discrimination based on America’s racist history.

Lastly, consider how different this rant would be if the interlocutor simply stated, “I am not concerned with the rights of the gay community and here’s why” and simply listed and described the social points. Instead, the interlocutor uses the repetition of *then I will worry about the gays* at the end of each and every point to emphasize their view of how little they care about the gay community.

#### 4.5 The Pejorative Noun *Illegal*

In the previous chapter I discussed Wierzbicka’s explanation of the change that occurs when adjectives are nominalized and a kind (or a category of entity) forms during this process. When this kind is formed (such as with *a blonde*), stereotypical properties become associated with it (such as +Dumb, +Sexy) for socio-cultural reasons. For *illegal*, I argue that Mexican is part of the prototypical idea of *illegal*. However, when the discourse topic changes, so too can

the prototype. To better understand the flexibility of the network of negative concepts that underline the prototype, terms that co-occurred with *illegal*, *poor*, *gay*, and *female* were collected and analyzed. If you look at Example 121, you will note that *terrorist* co-occurs with *illegal*.

[121] Hillary: Economic Plan: Keep the illegals coming, don't care if terrorists, we need their votes and too many dumb people vote for me anyway. (Twitter, Oct. 5, 2016)

While this seems to stray from the *Mexican* aspect of the prototype, it taps into a common narrative used in debates on immigration and homeland security. If the topic switches to terrorism, the Middle East, or Isis, the meaning of *illegals* can easily adapt its prototype to fit the current discourse. Example 122 below has another common narrative used in immigration discourse.

[122] It will now become “the party that gives away the most - wins” period wherein the taxpayers of America will support all the freebies our politicians will be promising the illegals and other leeches on society so they can get re-elected. (Commentary on washingtontimes.com article, 2015)

The co-occurrence of *leeches* with *illegal* brings to mind an image of a leech sucking the blood from some entity. This ties into a common metaphor of immigrants draining a country of valuable resources. As the cost of illegal immigration is a common discussion topic in immigration discourse, *leech* is a very common, negatively descriptive label attached to immigrants. Additionally, using the term *illegal* instead of *immigrant* or *undocumented* taps into the criminality element, bringing to mind images similar to Trump’s controversial comments about Mexican immigrants as seen below in 123:

[123] When Mexico sends its people, they’re not sending their best.... They’re bringing drugs. They’re bringing crime. They’re rapists. (Trump, June 16, 2015)

In Example 123, Trump is making an argument as to why undocumented immigrants must be kept out of the United States, stating that these immigrants are rapists and drug mules. This is a common narrative associated with immigration, especially when tied to Mexico or South America. However, when these claims about undocumented immigrants are contradicted, some still connect *criminal* to the term *illegal* as a reference to their status as in Example 124.

[124] That moment when you realize ALL illegals are technically criminals. (Reddit, 2016)

Again, this ties back to the negative adjectival meaning which has been amplified when it is used as a noun.

Example 125 is a Reddit conversation between multiple posters. The original poster proposed the return to the term *illegal alien* instead of *illegal immigrant*. The following example includes several reactions to this proposal.

[125] **S1:** Um, uhhh, um, excuse me. The correct term is “Not-Yet-Documented Opportunity Seeker.” Please don't be a xenobigot.  
**S2:** “Our new ethnic friends who haven't found the time to fill out some paperwork”  
**S3:** Or just call them illegals for short. That way you don't have to explain the difference between an alien and an immigrant.  
**S4:** Oh - they really get offended if you shorten illegal immigrants to just illegals  
**S5:** How can you DEHUMANIZE them like that. They're HUMAN BEINGS in the YEAR 2016 not illegal THINGS don't you have **FEELINGS**?  
**S6:** In fairness, it is dehumanization used as propaganda to make it easier for people who want them **OUT OUT OUT** to feel a little better about kicking millions of people back into a shitty country. SJWs usually aren't necessarily wrong about this stuff, they just prioritize their feelings over reality and reason. They also care way too much about phrasing things correctly considering they basically see these people as a statistic with which to flaunt their remarkable knowledge on social justice issues, or lack thereof really.  
**S7:** I disagree on the first part. “Illegals” is just calling it like you see it. Nobody feels good about deporting people, but that's exactly it - you have to separate the emotions. We all realize it sucks for illegals but that really just shouldn't

be America's problem and the future of our society is at stake.

First, there's a push from S3 and S7 to do away with the head nouns *immigrants* and *aliens* and simply shorten the term to *illegals*. In line 22, S7 explain that using *illegals* is *just calling it like you see it*. This is a very common argument used in favor of the term *illegals*, that shortening the term to the descriptive adjective focuses the attention on the 'problem,' in other words the illegal nature of their status, and away from immigration. Often the argument against the term points out that the adjective *illegal* cannot be applied to human beings, but rather their actions. While the opposition simply seems to equate actions and the human being as the same, directly applying the adjective *illegal* to the person itself. Consider a similar Reddit exchange in Example 126:

- [126] **S1:** Seriously, illegals ILLEGAL. The law was broken and while they may be nice people it doesn't entitle them to some special treatment.
- S2:** "No person is illegal."
- S3:** Except the ones who are illegal
- S4:** Okay, then go to any other country and just waltz wherever you want because no one is illegal. Oh wait you can't? Because there are borders to countries? What a novel concept.
- S5:** The person isn't illegal, their actions are.
- S6:** Because in their mind "A HUMAN CANNOT BE ILLEGAL REEEEEEEEEEEEEEEEE!!!!!!!!!!!!!!!"

This exchange has many examples of both sides of this common argument, though some are simply mimicked, such as in lines 3, 5-6, and 9. The crux of the argument has to do with what adjective *illegal* is modifying, a human being or an action. In line 1, S1 uses the second instance of *illegal* as a verb rather than an adjective, which seems to imply that *illegal* is an action rather than the human being. This is supported in the next sentence where it is reinforced



with *the law was broken*. However, in line 4 S3 reverts back to *illegal* modifying human beings, stating *except the ones that are illegal*.

#### 4.6 Conclusion

Through a closer, more detailed look at the data, each of the terms in question *female*, *gay*, *illegal*, and *poor* are shown to have specific characteristics in how they are used pejoratively. *Female* is more complex as it is already established as a noun, as well as it has other uses such as gender-based topics. *Poor* is most often used in a satirical or sarcastic manner to convey an individual or group's perceived attitude toward the poor. *Gay* can be used within the gay community without a pejorative meaning, but when people outside the community use *gay* it can carry a negative meaning. *Illegal* has a prototype tied to Mexico due to the current political discourse in which it is used, however this prototype can adapt based on the topic which allows *illegal* to co-occur with negative terms, such as *terrorist*.

Despite these unique aspects, all four adjective nominals are often used in an US vs. THEM type narrative, meaning the narrative is structured around two groups being contrasted and the speaker will likely have an allegiance to one group and speak negatively about the other group. For *female* there was *woman* vs. *female* and *man* vs. *female*. For *gay*, it would be those that object to homosexuality. For *illegal*, it would be those that object to undocumented immigrants within the United States. *Poor* is slightly different as it is used in a sarcastic manner, but even here, the person or people perceived to have the negative attitude would be the other side of the equation, such as Governor Paul LePage.

Van Dijk (1993) offers up an analysis of US and THEM narratives from a Critical Discourse Analysis (CDA) perspective. CDA is a specific approach within discourse analysis

that focuses on power and dominance and how that is manifested within a discourse. Van Dijk explains that CDA connects society, discourse, and social cognition. In society, there is social inequality through an imbalance between groups and classes, men and women, Whites and Blacks, the rich and the poor, as well as citizens and immigrants. CDA focuses on these inequalities and how the dominant groups' stereotypes, opinions, and views are expressed.

Van Dijk explains an US and THEM model is also commonly used within one group, such as Whites talking with Whites. In this model, the important parts are that positive aspects are attributed to the US (the group the speaker belongs to) and negative aspects to THEM (the group of the target forms) and that these expressions are based on the 'typical' THEM, using phrases such as *is always like that*. These comparisons function as a controlled polarization that taps into socially shared ideologies and stereotypes about a group. Van Dijk argues that the following are used as means to persuade within the US and THEM model:

- (a) Argumentation: the negative evaluation follows from the 'facts'
- (b) Rhetorical figures: hyperbolic enhancement of 'their' negative actions and 'our' positive actions; euphemisms, denials, understatements of 'our' negative actions
- (c) Lexical style: choice of words that imply negative (or positive) evaluations
- (d) Story telling: telling about negative events as personally experienced; giving plausible details about negative features of the events
- (e) Structural emphasis of 'their' negative actions, e.g. in headlines, leads, summaries, or other properties of text schemata (e.g. mentioning negative agents in prominent, topical position)
- (f) Quoting credible witnesses, sources or experts, e.g. in news reports (p. 264)

Many of the examples included in this chapter followed van Dijk's US and THEM model in one way or another. In justifying the use of *illegal*, speakers in Example 125 show a clear distinction between US vs. THEM when they discuss deporting immigrants (THEM) and stating that it is not America's problem (US). Similarly, *female* and *gay* are used with strong US vs. THEM distinctions, as seen in this chapter. More specifically to van Dijk's model, the

argumentation and rhetorical figures are employed when *illegals* are called *criminals* (124) and *rapists* (123), or when the term co-occurs with *terrorists* (121) and *leeches* (122). Likewise, lexical style can be seen with *female* in the pejorative compound *basic ass female* (109). Storytelling and anecdotal negativity is used in all four terms, but can be seen in this chapter when speakers state that all the gays they know are perverts (117). Van Dijk's model fits well with the data and only solidifies the pejorative uses of *illegal*, *gay*, *female*, and *poor* often follow an US and THEM narrative.

The hypotheses presented in this thesis now have support with my explanation of the clusters of linguistic properties in Chapter 3 and a detailed examination of the data presented in this chapter. In the next chapter, an empirical approach based in human judgment explores whether these hypotheses hold true when other people analyze the data.

## CHAPTER 5

### EMPIRICAL STUDY: HUMAN JUDGMENT<sup>11</sup>

#### 5.1 Introduction

In this chapter, I present results from two human judgment corpus studies. These two studies have different aims. In study 1, I collected annotations on a subset of the corpus from a group of experts with a linguistic background and calculate agreement within the group. The aim of this study is to validate my original annotations by having other linguists label a section of the same data I labeled, and then checking for agreement.

In study 2, we automatically extract additional data from Twitter which was annotated by a group of non-expert annotators through crowd-sourcing and agreement is calculated across the group. There are three aims for this study: (1) To see if nominal uses of *illegal*, *gay*, *female*, and *poor* occur in randomly extracted Twitter data. The originally compiled corpus was hand harvested specifically with the intent to study the pejorative forms. (2) To see whether nominal uses seem to be more pejorative than adjectival uses in the extracted Twitter data (3) and to see what extent non-expert annotators agree about the pejoration associated with these forms.

#### 5.2 Expert Judgment Study

Thus far the corpus only had my original annotations for pejoration. While these annotations gave a baseline for the study, it only revealed the perspective of one individual on the sentiment of *illegal*, *female*, *poor*, and *gay*. In an effort to balance out this limited perspective, I collected further annotations to compare with mine for agreement. I extracted a

---

<sup>11</sup> All tables in this chapter are reproduced from Palmer, A., Robinson, M., and Phillips, K. (2017). *Illegal* is not a noun: Linguistic form for detection of pejorative nominalizations. In *Proceedings of the First Workshop on Detecting Abusive Language Online* with permission from the Association for Computational Linguistics.

small section of the corpus (121 instances), now called SUB CORP. SUB CORP is spread over the four target forms, with 30 instances of *female*, 31 of *gay*, 34 of *illegal*, and 26 of *poor*, with 54 instances coded as pejorative and 67 as non-pejorative.

Two female and three male annotators were chosen who were either graduate students of linguistics or had recently graduated with their Master's in Linguistics. Along with my original annotations, this provided a gender balanced<sup>12</sup> set of six annotators with linguistic training.

The expert annotators were given annotation guidelines that were followed by the original annotator but were also told to trust their own judgment. These guidelines are discussed in Section 2.4 and are reproduced here:

- Negative adjective(s) modifying the target nominal form
- Co-occurrence with phrases referring to particular stereotypes or behaviors associated with the relevant referent group (e.g. *freeloading* with an occurrence of *poor*)
- Appearance near negative verbs such as *hate* or *despise*, or negative phrases such as *get rid of* or *hardly any good*
- Co-occurrence with other negative terms, such as *slut* for *female* or *wetback* for *illegal*
- Other negative implications not tied to a specific lexical item or phrase

The annotators were explicitly told to focus on the indicated grammatical form to determine pejoration and not to focus solely on the context. To illustrate this, consider Examples 127 and 128.

[127] Gay marriage is evil. (Constructed)

[128] A gay sat next to me on the bus. (Constructed)

While the context in example 127 is negative, the use of *gay* to modify marriage is not where the

---

<sup>12</sup> As the annotator's background could possibly influence judgment, future studies will strive to provide a demographic balance across annotators.

pejorative meaning is coming from. In example 128, there is nothing blatantly negative within the context, but depending on the annotator, the use of the indefinite singular of *gay* has a pejorative meaning. Therefore, an annotator who is following the guidelines will likely rate 128 as pejorative and 127 as non-pejorative. However, this proved to be a challenge for some of the annotators, especially with adjectival forms in a highly pejorative sentence.

### 5.2.1 Results

Table 5.1: Fleiss' Kappa Agreement between Expert Annotators

	<b>B2</b>	<b>B3</b>	<b>B4</b>	<b>B5</b>	<b>B6</b>
<b>A1</b>	0.461	0.703	0.691	0.747	0.551
<b>B2</b>		0.485	0.458	0.369	0.516
<b>B3</b>			0.642	0.586	0.483
<b>B4</b>				0.614	0.707
<b>B5</b>					0.483
<b>B6</b>					
Fleiss' Kappa (all)					0.561
Fleiss' Kappa (all except B2)					0.616
Fleiss' Kappa (A1, B3, B4, B5)					0.662

The top part of Table 5.1 shows the agreement between pairs of annotators measured with Cohen's kappa<sup>13</sup> (Cohen, 1960). The bottom part of the table shows the agreement across groups of annotators measured with Fleiss' kappa (Fleiss, 1971). A1 represents the original annotator, while B2-B6 are the added expert annotators.

While interpretations on kappa scores vary, Landis and Koch (1977) calculate a range of 0.41-0.60 as moderate agreement and a range of 0.61-0.80 as substantial agreement. The overall Fleiss' kappa score across annotators is 0.561. While this is only moderate agreement, it was

---

<sup>13</sup> Agreement for Cohen and Fleiss' kappa were computed in R using the IRR package.

discovered that annotators B2 and B6 were outliers, deviating from the other four annotators. Upon closer inspection of the annotations, it appears B2 and B6 struggled to focus solely on the grammatical form and not the context. B2 and B6 also tended to choose pejorative over non-pejorative labels overall in comparison to the other annotators. Comparing across all annotators with the exception for B2, there is substantial agreement with a Fleiss' kappa of 0.612. If B6 is removed as well, there is a kappa score of 0.662.

### 5.3 Crowd-Sourced Non-Expert Annotation

In addition to the expert agreement study, we wanted to see if the terms in question would still be judged as pejorative to annotators with no linguistic training in our second study. If this slight shift between adjective and noun indeed brings about a negative shift in meaning as well, speakers should be able to discern this shift using their own linguistic knowledge. With this in mind, we conducted a crowd-sourced empirical study using Amazon Mechanical Turk.

Amazon Mechanical Turk is a resource that provides researchers access to paid non-expert annotators. The benefit of such a resource is that a large body of annotation can be completed in a relatively short period of time and for a small amount of money. The workers, also called Turkers, can choose from a list of human intelligence tasks (HITs), offered up by researchers. Both the workers and the researchers remain anonymous throughout the process. The research indicates a pay rate for each task and the worker chooses which task they wish to complete. If the worker wishes, they can complete the series of tasks offered within a single study. The researcher then approves the work and releases the money to the worker.

The researcher has the option to restrict the tasks to workers with special qualifications, such as number of successfully completed tasks and whether previous pay was withheld due to

poor quality of work. Using these qualifications, we restricted the current corpus study to annotators that qualified as masters.

The value of non-expert crowd-sourced annotation varies over the task in question. According to (Snow, et al., 2008), this type of annotation combats the bias issue as it introduces a higher level of diversity than a single expert annotator or a small set of linguist annotators. For example, if we have 100 instances for annotation and collect 5 annotations per instance, the 500 annotations collected could come from 500 different annotators. Additionally, crowd-sourced annotation allows for a unique insight from non-experts on linguistic phenomena.

### 5.3.1 Methods

The original corpus (ORG CORP) was collected specifically with the goal of analyzing *illegal*, *poor*, *female*, and *gay* in pejorative use, so the corpus itself was initially unbalanced. To help counteract this, adjective and zero plural forms were added to the corpus (ADD CORP) to provide a more balanced corpus in terms of pejorative and non-pejorative forms.

To better test the hypothesis that certain adjectives have a pejorative meaning when used as nouns, a third corpus (AUTO CORP) of nominal forms of *illegal*, *gay*, *female* and *poor* was automatically extracted using *twarc*<sup>14</sup>. *Twarc* is a Python toolkit for processing Twitter data from within a 2000-mile radius of the geographic center of the United States. 100,000 tweets were extracted from Twitter with one of the target forms in each instance. Next, the tweets were filtered and only tweets with six words or more were kept, not including the username, hashtags, emoticons, URLs, numbers, or punctuation, which left us with a total of 56,000 tweets.

---

<sup>14</sup> <https://github.com/DocNow/twarc>



A part of speech (POS) tagger was then used to isolate the desired target forms. During this process, we ran into issues specifically with the indefinite (*a gay*) and definite singular forms (*the illegal*). Consistently, the POS tagger tagged these singular forms as adjectives rather than nouns. Therefore, we constructed a filter to combat this issue called the NomCatcher (Palmer et al., 2017). The filter corrected this POS labelling error by basically stating that if something that looks like an adjective but is followed by a verb or punctuation to indicate the end of a sentence, label it as a noun.

[129] The illegal race is exciting to watch. (Constructed)

[130] The illegal raced down the street. (Constructed)

The NomCatcher would label *illegal* in Example 129 as an adjective, as it is followed by the noun *race*, but in Example 130 the NomCatcher would more accurately label *illegal* as a noun and instead of an adjective because it is followed by the verb *raced*.

Using the NomCatcher, the target forms were labeled with the appropriate grammatical form. From these labeled instances, 200 instances were randomly chosen for each target form, 100 were adjective forms and 100 nominals, with a total of 800 instances overall. The instances were shuffled and randomly compiled into five batches of 40 in each target form, totaling 20 batches. Additionally, 200 instances were chosen at random from the ORG CORP were also added to do a comparison against the original annotations.

Five Turkers annotated each batch and were paid \$0.50 per batch. The annotators were given the instructions to label the highlighted forms as either positive, neutral, or negative. These labels were chosen over the original pejorative and non-pejorative labels simply for ease of understanding. In this schema, positive and neutral labels would equate to non-pejorative

whereas the negative label would equate pejorative. To ensure their understanding of the task, the following constructed examples were given in the instructions:

a POSITIVE: If you want the job done right, ask *a female* to do it.

b NEGATIVE: I don't understand why *females* think they know how to drive.

c NEUTRAL: My first pet ever was a *female* lizard.

Annotators were reminded to focus on the sentiment of the underlined word itself and not necessarily the context as a whole. Again, this point appeared to be difficult for some annotators to accomplish, similar to the struggles of the expert linguist annotators B2 and B6. Consider Example 131:

[131] I am anti-gay because I feel gays and lesbian deserve a better life then one that seems to be affected by depression, suicide, promiscuity, domestic violence, and alcohol abuse to name just a few things.  
(blacknright.wordpress.com, 2009)

The Turkers annotated the example above twice, once for the adjective form of *gay* in *anti-gay* and once for the bare plural *gays*. Both are examples of annotations that proved to be difficult for our crowd-sourced non-experts.

For the adjectival form, while it is in the negative context of *anti-gay*, there is nothing inherently pejorative about the adjectival part. The negative meaning comes with the meaning of *anti*, yet it is understandable why some annotators would struggle to label this as non-pejorative.

Likewise, with the bare plural form, the interlocutor is expressing the desire for a better life for *gays*, which can be seen as a positive. However, this is contrasted with a list of negative aspects that are being associated with *gay*, such as *suicide*, *promiscuity* and *domestic violence*. The point being, some annotators will see the context of *a better life* and rate the bare plural as

positive while other annotators will read the implications associated with this instance and label if pejorative.

### 5.3.2 Results

Table 5.2 displays agreement based on majority vote over the five annotators that labeled each instance.

Table 5.2: Agreement between Annotators via Majority Vote

n=800	<b>5agree</b>		<b>4agree</b>		<b>3agree</b>		<b>NoMaj</b>	
	14.5%		26.3%		46.9%		12.3%	
<b>Sent.Label</b>	N	A	N	A	N	A	N	A
NEG	72	16	61	51	71	67		
NEUT	11	8	34	49	83	120		
POS	3	6	4	12	14	20		
							47	51

The percentages indicate how often that level of majority agreement was reached over the instances. In each column the N stands for noun forms and the A stands for adjective forms. Below this, the numbers are raw numbers indicating how many nouns or adjectives were labeled for each category. 5agree, 4agree, and 3agree indicate how many annotators agree over five annotators. At least three annotators are needed for a majority vote. NoMaj stands for no majority, or less than three annotators that agree per batch.

Majority agreement (3agree-5agree) was about 87% out of the 800 instances and complete agreement (5agree) reached about 15%. While 15% might seem insignificant, interestingly of these complete agreements, the majority of the instances were nouns and were labeled negative. Meaning, while complete agreement is understandably difficult to reach, labeling nouns as negative was very salient. This shows non-expert annotators can more readily agree that when *poor*, *female*, *illegal*, and *gay* are used as nouns the meaning is pejorative.

The areas of disagreement among the annotators tended to be in the use of positive as a label as well when the instance contained an adjective. Overall, the non-expert annotators were more consistent when labeling nouns.

Next, we analyzed label agreement with grammatical form as seen in Table 5.3.

Table 5.3: Correlation between Grammatical Form and Sentiment

<b>Majority Vote</b>	<b>Adj</b> 400	<b>Noun</b> 400
%NEGATIVE	33.5	<b>51.0</b>
%NEUTRAL	<b>44.2</b>	32.0
%POSITIVE	<b>9.5</b>	5.2
%NOMAJ	12.8	11.8

Fifty-four percent of adjectival forms were labeled as non-pejorative (positive and neutral) at a majority level (3 agree or more). Fifty-one percent of nominal instances were labeled as pejorative, and 37% were labeled non-pejorative.

### 5.3.3 Non-Expert Annotation against Original Expert Annotation

During the crowd-sourcing annotation, 10 instances from the ORG CORP were randomly added to each batch. While this number is too small to make any major assessment, the comparison is still noteworthy. Overall, 200 instances from the original corpus were annotated by the crowd-sourcing annotators. Reminder, the labels used by the original annotator (Pej, Non-Pej, Unc, Sat) were different from what the crowd-sourcing annotators used (Pos, Neg, Neut.)

In Table 5.4, the Turkers' annotations of 200 instances from ORG CORP are compared against my original annotations. The results are in raw numbers, but percentages across the original label are included as well. The labels on the left-hand side are my original labels and the labels across the top are the labels the Turkers used. NOAN indicates two instances that were not annotated at the time. Of the instances labeled pejorative, the crowd-sourcing annotators mostly

labeled them as negative (69%), of the non-pejorative instances the labels were split between negative (33%) and neutral (49%). The satire/sarcastic instances were mostly labeled as negative (59%), and the majority labeled unclear instances as neutral (61%) or did not reach a majority agreement (28%).

Table 5.4: Crowd-Sourced Annotations Compared against Original Annotator

Expert	CS: Neg	CS: Neut	CS: Pos	NoMaj
PEJ	86 (69%)	26 (21%)	1 (1%)	12 (9%)
NONP	11 (33%)	16 (49%)	2 (6%)	4 (12%)
SAT	13 (59%)	4 (18%)	1 (5%)	4 (18%)
UNC	2 (11%)	11 (61%)	0	5 (28%)
NOAN	2 (100%)	0	0	0

#### 5.4 Conclusions

The empirical studies done with both expert and non-expert annotators provided encouraging support to the hypotheses argued in this thesis. According to human judgment, there does appear to be a correlation between pejorative meaning and the nominal forms of the adjectives in question. Additionally, while humans tend to have a natural bias in language judgment, this can be balanced with a pool of annotators to provide a more reliable result. Finally, human judgment is a valuable and successful tool in analyzing subtle semantic shifts that might not be uncovered through other approaches. Meaning within language is complex and multifaceted, which makes it a challenge to analyze. However, humans have an ability to analyze the various layers of meaning with language that is difficult to replicate through other approaches.

In an effort to further test these hypotheses, the next chapter presents a study using a computational approach to language judgment using sentiment analysis.

## CHAPTER 6

### SENTIMENT ANALYSIS

#### 6.1 Introduction

Thus far, I have discussed the data in such terms as pejorative and non-pejorative, or negative, neutral, or positive. When discussing negative meaning or positive meaning, what is really being discussed is the sentiment of a word or phrase. Sentiment relates to negative or positive feelings or opinions about something. Therefore, in analyzing terms for pejorative meaning, one might consider approaches to analyzing sentiment other than human judgment. In this case, we investigate the effectiveness of computational methods for automatic analysis of sentiment.

As complex as language and meaning can be, one might wonder why a computational approach would be suitable for this study. In fact, Bender (2008) argues that computational methods should be used to test linguistic hypotheses. She explains this is for two reasons: (1) languages are made up of multifaceted subsystems and computational methods can account for these subsystems, while human analysis tends to focus on one subsystem at a time. (2) Computational methods allow for a larger amount of data to be analyzed in a timely manner than what is possible through human analysis. Additionally, if the sentiment analysis methods provide an accurate analysis of the data, this would open up the potential of analyzing much more data for future empirical studies.

With these goals in mind, I present in this chapter the methods and results of a study of what can be learned by applying automated sentiment analysis methods to the data I have collected.

## 6.2 Methods

Sentiment analysis is a computational method that seeks to determine the overall intended sentiment of a text<sup>15</sup>. Sentiment analysis is most often used to analyze campaign success on social media, product reviews, movie reviews, as well as monitor public relation problems. For example, if a company releases an ad and wants to know how people are reacting to the ad, sentiment analysis can be used to examine attitudes expressed when the ad is mentioned on social media. Likewise, if a hotel wanted to pinpoint commonly mentioned aspects in negative reviews, such as ‘dirty towels’, sentiment analysis could be implemented.

The semantic shifts seen in *illegal*, *poor*, *gay*, and *female* are both subtle and more complex than the types of sentiment associated with hotel reviews (for example), making it an open question whether the standard sentiment analysis methods will be suitable for addressing this data. As this was a first step toward investigating computational methods for this analysis, we deploy a rather straightforward approach to sentiment. The reasoning for this was twofold, (1) to see what we can learn through simple methods, and (2) be able to adjust our methods based on what we learn. Specifically, we use a lexicon-based approach to sentiment analysis. This approach utilizes a large list of words that are annotated with a number that corresponds with either a positive, neutral, or negative meaning. For example, the positive adjective *pretty* might have a score of +1, the neutral adjective *average* might have a score of 0, and the negative adjective *ugly* might have a score of -1. These scores are based on annotated lexicon lists.

Using the lexicon or lexicons chosen, words that bear sentiment in a sentence or text will be given a sentiment score as seen in Figure 6.1.

---

<sup>15</sup> For a detailed survey of methods for sentiment analysis and opinion mining, please see: Pang, B., & Lee, L. (2018). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.

0
-1
T  
 America is being menaced by gays!

**Sentiment Score: -1**

Figure 6.1: Sentiment Analysis Sentence Example

If the content words are included in the lexicon, they are marked with a sentiment score and then a total score will be added up for the entire sentence. The target word itself receives no score. In the lexicon approach, a positive word (+1) in a sentence will cancel out a negative word (-1) and the result would be a neutral reading, as seen in Figure 6.2.

+1
0
0
0
0
T
-1  
 As a proud member of the middle class, I can't wait for all of the poors to die off.

**Whole sentence: 0**

Figure 6.2: Sentiment Analysis Neutral Rating

### 6.3 Data

The lexicon sentiment analysis approach was applied to the ORG CORP and SUB CORP, which were described in Chapter 2 and Section 5.2. The addition of the SUB CORP allows for a more balanced corpus as it includes nominal, adjective, and zero plural forms spread over pejorative and non-pejorative instances as seen in Table 6.1.

Table 6.1: Sentiment Category Breakdown of *Female*, *Gay*, *Poor*, and *Illegal*

	Female	Gay	Poor	Illegal
Pejorative	521	101	59	457
Non-Pejorative	472	356	165	482
Satire/Sarcasm	1	13	252	22
Unclear	116	21	21	74
	1,110	491	497	1035



Using the ORG CORP and SUB CORP, a total 3133 instances were included in this computational study.

#### 6.4 Lexicons

Before the computational study was conducted, several annotated lexicons were considered and tested to find options suitable for the type of language data in the corpus. As sentiment analysis methods are used with various different purposes, from analyzing product reviews to assessing the attitudes to political candidates on social media, the type of language in each lexicon is different. Some lexicons are more adept to social media language, including common hashtags and internet-based slang acronyms such as ROFL. Additionally, some lexicons focus on a more fine-grained approach to affect rating that goes beyond just positive, negative, and neutral.

These more detailed lexicons include categories such as arousal and dominance. Some lexicons also include subtle sentiment scores for words such as *well*, *like*, *just*, and *pretty*, scoring these words as positive. While each of these words does have a positive meaning, in the present corpus these words are more likely to be used as discourse markers. Likewise, words such as *tall*, *shower*, *days*, and *walk* were scored as positive, while *ergo*, *day*, *name*, and *so* were scored as negative.

To get a clear idea what type of lexicon would be best for our data, I evaluated 8 lexicons: MPQA Subjectivity Lexicon (Wilson et al., 2005), Harvard General Inquirer (Stone et al., 1966), Bing Liu's Opinion Lexicon (Liu et al., 2005), Sentiment Composition Lexicon of Opposing Polarity Phrases (Kiritchenko and Mohammad, 2016), SemEval-2016 Task 7-Determining Sentiment Intensity of English and Arabic Phrases (Kiritchenko et al., 2016), Macquarie

Semantic Orientation Lexicon (Mohammad et al., 2009), Warriner-Kuperman Affective Ratings (Warriner et al, 2013), and SentiWordNet (Baccianella et al., 2010).

For this evaluation process, 61 instances spread over the four target forms were randomly extracted from the corpus to use as a test suite. All 8 lexicons were then used to score the 61 instances. I then inspected the details of the scoring to see which words were scored, which words were not. After analyzing the scorings, I noticed some lexicons had scored words such as *day*, *name*, *mind*, and *ergo* as negative and *call*, *early*, *put*, and *shower* as positive. As this level of aspect did not suit our data, these lexicons were excluded. In the end, MPQA, Harvard General Inquirer, and Bing Liu's Opinion Lexicon were chosen as the words included as well as the sentiment rating were straightforward and fit the needs of the study. These lexicons are described below.

#### 6.4.1 MPQA Subjectivity Lexicon

The MPQA Subjectivity Lexicon (Wilson et al., 2005) has a total of 8,222 words. The lexicon is part of OpinionFinder, a system that automatically identifies subjective sentences and the sentiment of a text. The Subjectivity Lexicon provides clues for possible subjectivity in the form of a list of words. This lexicon was compiled from various other projects and is a mix of manually annotated and automatically annotated sources. Each word on the list is annotated for part of speech, level of subjectivity (either weak or strong), and labeled as negative, positive, or neutral. Some example sentiment scorings from MPQA: *abusive*-negative, *signal*-neutral, *persuasive*-positive.

#### 6.4.2 Harvard General Inquirer

The Harvard General Inquirer lexicon (Stone et al., 1966) has a total of 11,788 words. The lexicon was compiled from the Harvard and Lasswell general-purpose dictionaries. Each word is labeled as negative or positive in the lexicon, and additional categories of sentiment are included such as hostile, pleasure, pain, feel, arousal, virtue, and vice. Some example sentiment scorings from Harvard General Inquirer: *questionable*-negative, *discreet*-positive, *needy*-negative, *flashy*-positive.

#### 6.4.3 Bing Liu Opinion Lexicon

Bing Liu's Opinion Lexicon (Liu et al., 2005) contains 6,800 words. Words in the lexicon are labeled as positive or negative, with neutral not included in the lexicon. This lexicon includes common misspellings, morphological variations, and slang words. A lexicon that includes common misspellings is especially useful with online data as there are frequent accidental and creative spellings, such as *benifits* and *f\*\*k*. Some example sentiment scorings from the Opinion Lexicon: *strong*-positive, *lengthy*-negative, *accomplish*-positive, *garbage*-negative.

### 6.5 Scoring Methods and Evaluation

For our analysis, each label was converted into a numeric score using Python: positive (+1), negative (-1), and neutral (0). For example, if a lexicon labels the word *happy* as positive, the score for this word would be converted to +1. As described above, the word level scores are combined to compute a sentence-level score. At the sentence level, we created four possible scoring categories: positive, negative, conflicting neutral, and neutral. The neutral category was

split into two separate categories to differentiate between an instance that was neutral due to sentiment bearing words canceling each other out and an instance where no words were scored for sentiment.

The scores of each word are added up and once each instance has a final score the whole instance is categorized as either positive, negative, conflicting neutral, or neutral. This final score is compared against the author's earlier annotation. The labels used in the earlier annotation were pejorative, non-pejorative, sarcasm/satire, and unclear. For non-pejorative, both positive and neutral labels were included. For sarcasm/satire, we anticipated a pejorative score as the type of sarcasm seen in the corpus were not the standard positive sentiment bearing examples.

## 6.6 Results

The following tables show the results of the whole instance sentiment analysis approach, broken down by lexicon used and the target form. The labels along the left side indicate my sentiment annotation, while the sentiment labels along the top are the labels produced through sentiment analysis.

As can be seen, the sentiment analysis scoring for *female* does not reach very high agreement with my sentiment annotation. Of the instances I labeled as pejorative, the Bing lexicon scored about 38% of them as negative, with Harvard General Inquirer and MPQA slightly less with 32% and 28% respectively. Notice also that 22% (Bing) and 42% (Harvard) of those pejorative instances were scored as neutral, meaning none of the words in the instance were scored as having sentiment. These non-scoring neutral instances could be one explanation for the low agreement results.

Table 6.2: Sentiment Analysis Results for *Female*

Female-1108 Instances	Bing				
	Pos	Neg	Confl Neut	Neut	Total
Pejorative	142	200	66	112	520
Non-Pejorative	154	153	46	118	471
Sarcastic/Satire	0	0	0	1	1
Unclear	45	18	15	0	107
	Harvard				
	Pos	Neg	Confl Neut	Neut	Total
Pejorative	104	164	32	220	520
Non-Pejorative	124	101	34	212	471
Sarcastic/Satire	0	0	0	1	1
Unclear	27	18	2	60	107
	MPQA				
	Pos	Neg	Confl Neut	Neut	Total
Pejorative	230	143	97	50	520
Non-Pejorative	223	138	47	63	471
Sarcastic/Satire	1	0	0	0	1
Unclear	53	23	11	20	107

Table 6.3: Whole Instance Sentiment Analysis Results for *Gay*

Gay-491 Instances	Bing				
	Pos	Neg	Confl Neut	Neut	Total
Pejorative	26	44	10	21	101
Non-Pejorative	111	116	24	107	358
Sarcastic/Satire	4	7	0	2	13
Unclear	3	4	1	5	13
	Harvard				
	Pos	Neg	Confl Neut	Neut	Total
Pejorative	23	17	10	51	101
Non-Pejorative	110	68	37	143	358
Sarcastic/Satire	1	2	1	9	13
Unclear	3	2	0	8	13
	MPQA				
	Pos	Neg	Confl Neut	Neut	Total
Pejorative	45	29	19	8	101
Non-Pejorative	168	103	48	39	358
Sarcastic/Satire	5	6	1	1	13
Unclear	6	1	3	3	13

A similar pattern of results can be seen with *gay*. Once again, Bing does slightly better than the other two lexicons, but the agreement is not high. Of the pejorative instances, only 44% were scored as negative by Bing, 17% by Harvard, and 29% by MPQA. Again, the neutral category shows that a large percentage of these instances are receiving no sentiment score at all. 51% of the pejorative instances were non-scoring neutral instances with the Harvard General Inquirer lexicon. Additionally, it is important to note that *gay* has a smaller number of instances (101 pejorative).

Table 6.4: Whole Instance Sentiment Analysis Results for *Poor*

Poor-376 Instances	Bing				
	Pos	Neg	Confl Neut	Neut	Total
Pejorative	6	11	6	14	37
Non-Pejorative	31	58	19	34	142
Sarcastic/Satire	50	43	32	56	181
Unclear	4	3	1	5	13
	Harvard				
Pejorative	3	11	1	22	37
Non-Pejorative	25	37	8	72	142
Sarcastic/Satire	30	34	10	107	181
Unclear	3	1	0	9	13
	MPQA				
Pejorative	11	9	6	11	37
Non-Pejorative	51	43	35	13	142
Sarcastic/Satire	84	37	24	36	181
Unclear	5	2	1	5	13

With *poor*, the sarcastic/satire category is the important one to consider, as the label was used much more than pejorative. In Section 2.4, I discussed the usual way sarcasm functions. Suli (2016) argues that positive words exist in sarcastic utterance and the implication is the exact opposite of the positive context. Therefore, in sentiment analysis, sarcastic instances should have a higher positive rating. However, in Section 2.4, I also discussed how the sarcasm/satire instances in the present corpus behave differently as the intent is to voice a negative perception.

Despite this, there is not a huge difference between the positive or the negative scores for the sarcastic/satire instances. For both Bing and Harvard, the positive and negative scores are somewhat balanced with 24-28% range and 17%-19% range respectively. Only MPQA seems to favor the positive label with 46%. Once more the non-scoring neutral category contains a large portion of the sarcastic/satire instances, with Harvard having 59% of the instances with no score.

Table 6.5: Whole Instance Sentiment Analysis Results for *Illegal*

Illegal-1034 Instances	Bing				
	Pos	Neg	Confl Neut	Neut	Total
Pejorative	91	153	46	166	456
Non-Pejorative	99	129	49	205	482
Sarcastic/Satire	9	4	3	6	22
Unclear	11	12	3	1	48
	Harvard				
Pejorative	72	83	15	286	456
Non-Pejorative	77	116	24	265	482
Sarcastic/Satire	4	6	0	12	22
Unclear	3	5	0	40	48
	MPQA				
Pejorative	160	120	64	112	456
Non-Pejorative	169	120	41	152	482
Sarcastic/Satire	11	6	4	1	22
Unclear	15	11	3	19	48

Again, with *illegal* the results are similar with a low percentage of the pejorative instances scored as negative and a considerable number receiving no score at all. Bing does slightly better than the other two lexicons with 34% scored as negative, but Bing also gives 36% of pejorative instances no score at all. Overall, the lexicons seem to have a general balance between positive scores and negative scores for the pejorative instances, with Bing slightly leaning toward negative and MPQA slightly leaning toward positive. These split, mixed results seem to indicate using sentiment analysis methods on this data will not give us a clear, consistent insight of the sentiment.

## 6.7 Discussion

Overall, the results of the sentiment analysis are conflicting and inconsistent. Lexicon-based sentiment analysis relies entirely on the context of the instance, meaning if the context is negative, the system will score instance as negative. It is reasonable to expect that this strictly context-driven approach might fail to capture some instances when pejorative meaning actually occurs within the grammatical form, as predicted by my hypothesis. In other words, *illegal*, *gay*, *female*, and *poor* develop their pejorative meaning in large part by being used as nouns, and that semantic shift within the grammatical form cannot be accounted for through sentiment analysis. This is not to say the adjective nominals do not occur in negative contexts, but simply that their pejorative meaning is not dependent on the forms being in a negative context. The fact that human annotators seemed to agree that grammatical form correlates with sentiment, and the computational approach does not, only further supports the hypothesis that the pejorative meaning is more strongly tied to the forms being nouns than to being in a negative context.

Consider the two examples in Figure 6.3 from the corpus and how each lexicon scored them.

*There are barely any good females these days.*

Bing: +1 [good +1]

Harvard: 0 [Scored no words]

MPQA: 0 [barely -1, good +1]

*But would you vote for a female who you believed had the qualities to be the president of the United States of America?*

Bing: 0-scored no words

Harvard: 0-scored no words

MPQA: 0-scored no words

Figure 6.3: Examples of Sentiment Scoring Over Lexicons

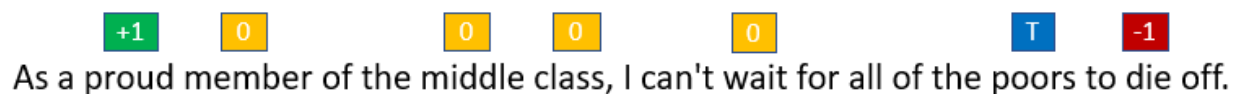
I labeled both instances as pejorative, though they were both considered neutral save for Bing



rating the first instance as positive. In our methods of sentiment analysis, each instance is treated as a list of words. The implications in instances above can be perceived by human annotators, but not through sentiment analysis.

While the results of lexicon-based sentiment analysis methods do not correlate well with the pejoration annotations for this data, there are other limitations that could have contributed to the results. Some of these limitations can be corrected with different methods and might produce slightly more accurate results in future studies.

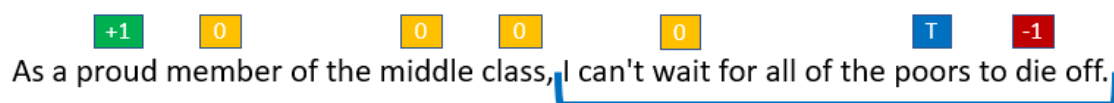
One such limitation is the approach of analyzing the entire instance rather than looking at the clause or phrase the target form occurs in. Consider Figure 6.4:



Whole sentence: 0

Figure 6.4: Whole Rating Example

Based on the position of the target word in the sentence in Figure 6.4, analyzing the entire sentence gives a false neutral score. While the system scored *die* as negative, it also scores *proud* as positive, though this positive modifier has nothing to do with the target form *poors*. Instead, the analysis could be restricted to the clause the target exists in to get a better analysis of the sentiment relating to the target form, as seen in Figure 6.5.



Whole sentence: 0

Clausal: -1

Figure 6.5: Sentiment Analysis Isolated Clause Approach

An additional limitation with the current approach is that it simply treats each instance as a list of words and does not consider word relationships and dependency. It stands to reason that direct modification of the target form, such as *fat female*, should be given more weight due to the direct relationship of sentiment and target word. In this approach, a prenominal negative modifier produces a negative score and a prenominal positive modifier will produce a positive score, as seen in Figure 6.6. Additional common dependency structures should also be considered.

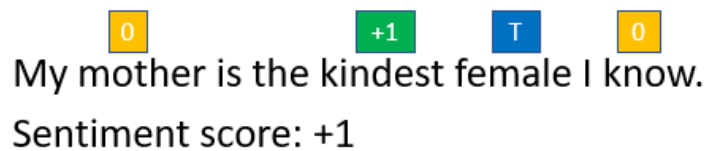


Figure 6.6: Sentiment Analysis Positive Adjective

Now comparing these different approaches shows how much variation can occur and how the limited results of our first attempt could be skewed, as shown in Figure 6.7. Analyzing the whole instance results in a neutral score of 0, but the clausal and prenominal methods are in agreement with a negative score of -1.

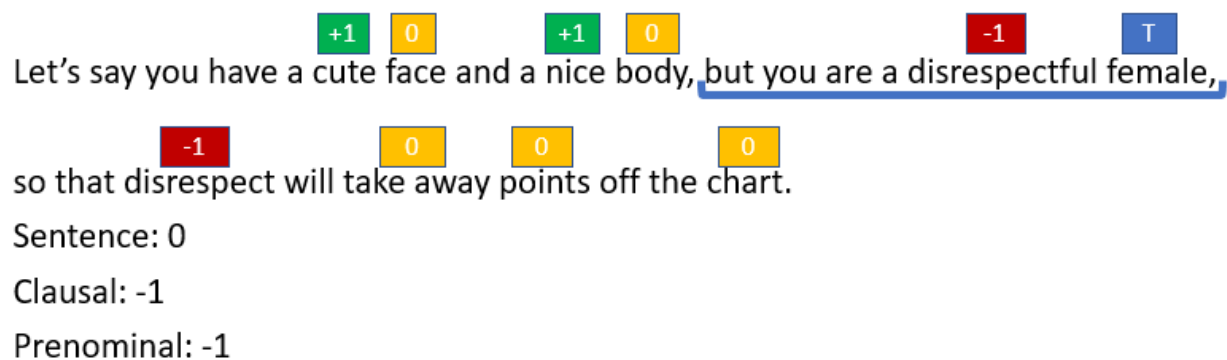


Figure 6.7: Sentiment Analysis Three Approaches

Beyond these limitations, there are some that cannot be accounted for with new methods, such as implications that go beyond sentiment bearing words. Consider Example 132:

[132] A man needs a female like a fish needs a lobotomy. (YouTube commentary, 2016)

For a human annotator, the implications of this statement are easier to understand. First, the annotator is likely familiar with the overall structure: an *A* needs a *B* like a *C* needs a *D*. This structure familiarity will already leave the overall implication that *A* does not need *B* even if the comparison is not understood. Additionally, there is an element of absurdity that comes with this familiar form. Here, the statement *a fish needs a lobotomy* is absurd, and the fact that it is absurd implies that *a man needing a female* is absurd as well.

However, when the sentence is analyzed using the lexicon approach, the implications of the structure and absurdism is completely lost. Figure 6.8 shows that when the words are tagged as neutral the overall score is neutral and the overall meaning is not properly captured.

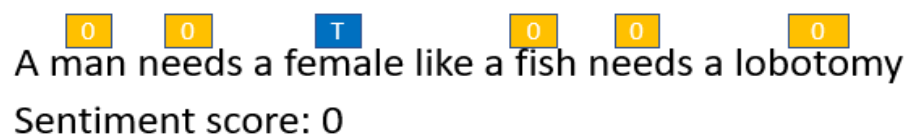


Figure 6.8: Sentiment Analysis False Neutral

In the end, future sentiment analysis experiments could be conducted on the same data, addressing at least some of the limitations discussed above.

## 6.8 Conclusion

The results from the sentiment analysis study support the hypothesis that *illegal*, *gay*, *poor*, and *female* are pejorative because they are nominalized. The results substantiate that the negative sentiment that our human annotators perceive is not restricted to the context of the sentence, but rather is conveyed through the grammatical form. Despite the fact that this was not the initial goal for our sentiment analysis approach, it nonetheless provided a means to corroborate the results of the human judgment studies.

Furthermore, computational approaches might still provide stronger support in future research if the limitations discussed above are taken into consideration. With adjusted methods, such as restricting the analysis to the clause level, a sentiment analysis approach might produce results that better resemble how human annotators judge this data.

## CHAPTER 7

### CONCLUSION

#### 7.1 Discussion and Implications

In this thesis, I have covered the various phenomena involved the pejorative meaning of *illegal*, *gay*, *poor*, and *female*. In Chapter 3, I discussed Wierzbicka's explanation that when an adjective is used as a noun, a kind of entity is created, and stereotypical properties become associated with that kind. Likewise, I showed that it is not simply nominalization that led to the negative meaning of *illegal*, *poor*, *gay*, and *female*; instead the nominalization of these adjectives led to a cluster of linguistic properties that contributed to the overall pejorative meaning. The role of definiteness and indefiniteness, specific and generic reference, as well as stereotypes and prototypes contribute to making an environment that is conducive to pejorative meaning.

I also presented empirical studies that support what was discussed in Chapter 3. The studies in Chapter 5 show that there a tendency for humans to perceive *illegal*, *gay*, *poor*, and *female* as pejorative when used as nouns. In Chapter 6, the sentiment analysis study supports the idea that the pejorative meaning our human annotators perceived was not the result of a negative context but the fact that *illegal*, *gay*, *poor*, and *female* were nouns. These studies suggest that there should be further investigations into these and other adjectival nominalizations.

#### 7.2 Future Research

There are many different directions future research could go with this topic. One possible future study is to look at other adjectives that follow a similar pattern, both pejorative and non-pejorative. This type of adjectival nominalization is a productive process that can be applied to many other adjectives beyond *illegal*, *poor*, *gay*, and *female*. This adjectival nominalization is

used often in American English, both in pejorative and non-pejorative ways. Some examples that were not investigated in this thesis are *undocumented*s, *deplorables*, *normals*, *abnormals*, *lames*, *grown*s, *young*s, and *old*s. Each of these examples seem to follow a similar process of converting an adjective into a noun; however, the degree of negative meaning likely varies over the terms. This research conceivably can be a starting point for further studies in other possible innovative forms.

My personal areas of interests for future studies are more in the productive process and the semantic shift that occurs. Therefore, a future study I hope to pursue is one that analyzes the degrees of pejorative meaning as they vary across grammatical forms. For example, do people perceive the indefinite singular (*a gay*) as more pejorative than the bare plural (*gays*)?

Another area of study I would like to continue to pursue is modifying methods for another sentiment analysis study to see if there is something further to learn. Furthermore, I would like to explore other computational methods to find if something else is better suited to analyze the subtle nature of this sentiment. I feel that some form of computation approach could strengthen our understanding of this pejorative process.

APPENDIX  
GRAMMATICAL FORMS

Grammatical Forms		Reference	Examples
Indefinite singular	an + noun	Generic and non-generic	<i>a gay is</i>
Definite singular	the + noun	Generic and non-generic	<i>the gay is</i>
Bare plural	∅ + noun + -s	Generic only	<i>gays are</i>
Definite plural	the + noun + -s	Generic and non-generic	<i>the gays are</i>
Zero plural	the + noun + ∅	Generic only	<i>the gay are</i>



## REFERENCES

- Acton, E. K. (2014). *Pragmatics and the social meaning of determiners* (Doctoral dissertation). Stanford University, Stanford, CA.
- Alim, S.H. and Smitherman, G. (2012). *Articulate while Black*. Oxford: Oxford University Press.
- Baccianella, S., Esuli, A., and Sebastiani, F. (2010) SentiWordNet 3.0: An enhanced lexicon resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, pages 2200-2204, Valletta.
- Bender, E.M. (2008). Grammar engineering for linguistic hypothesis testing. In N. Gaylord, A. Palmer, & E. Ponvert (Eds.), *Proceedings of the Texas Linguistics Society X conference: Computational linguistics for less-studied languages* (pp. 16-36). Stanford: CSLI Publications. Retrieved from [http://csli-publications.stanford.edu/TLS/TLS10-2006/TLS10\\_Bender.pdf](http://csli-publications.stanford.edu/TLS/TLS10-2006/TLS10_Bender.pdf).
- Bolinger, D. (1980) *Language, the loaded weapon: The use and abuse of language today*. London: Longman Group Limited.
- Cameron, D. (2016, April 3). To gender or not to gender? (Thoughts prompted by the death of Zaha Hadid) [Web log post]. Retrieved from <https://debuk.wordpress.com/2016/04/03/to-gender-or-not-to-gender-thoughts-prompted-by-the-death-of-zaha-hadid/>
- Carlson, G.N., & Pelletier, F.J. (Eds.). (1995). *The generic book*. Chicago: University of Chicago Press.
- [CNN]. (2015, April 21) *Female CEO: Women should not be president* [Video File]. Retrieve from <https://www.youtube.com/watch?v=myhdQciFsH0>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46. Retrieved from <http://journals.sagepub.com/doi/pdf/10.1177/001316446002000104>
- Croft, W. (1991). *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press.
- Dinan, S. (February 25, 2015). Civil rights commissioner warns of illegals voting under Obama amnesty. *The Washington Times*. Retrieved from <https://www.washington-times.com/news/2015/feb/25/civil-rights-commissioner-warns-illegals-voting/#.VO4Kd0r7GtM.twitter>
- Elite, T. [macklessons]. (2010, April 22) *Episode #235- "Basic ass females"* Retrieved from <http://macklessonsradio.com/index/episode-235-basic-ass-females->

- Female, n. and adj. (2017) *OED Online*, Oxford University Press, Retrieved from [www.oed.com/view/Entry/69157](http://www.oed.com/view/Entry/69157)
- Fleiss, J. L. (1971) Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76 (5), 378–382. Retrieved from <http://dx.doi.org/10.1037/h0031619>
- Galinsky, A.D., Wang, C.S., Whitson, J.A., Anicich, E.M., Kurt, Hugenberg, K., and Bodenhausen, G.V. (2013). The Reappropriation of stigmatizing labels: The Reciprocal relationship between power and self-labeling. *Psychological Science*. 24(10), 2020-2029. Retrieved from <http://faculty.wcas.northwestern.edu/bodenhausen/reapp.pdf>
- Gay, adj., adv., and n. (2017) *OED Online*, Oxford University Press, Retrieved from [www.oed.com/view/Entry/77207](http://www.oed.com/view/Entry/77207).
- Günther, C. to appear. The rich, the poor, the obvious – Arguing for an ellipsis approach to “adjectives used as nouns”. In A. H.-C. Leung & W. van der Wurff (eds.), *NPs in English: past and present*. Amsterdam: John Benjamins. Retrieved from [http://www.anglistik.hhu.de/fileadmin/redaktion/Fakultaeten/Philosophische\\_Fakultaet/Anglistik\\_und\\_Amerikanistik/Ang3\\_Linguistics/Dateien/Detailseiten/Guenther/Guenther\\_2015\\_The\\_rich\\_the\\_poor\\_the\\_obvious.pdf](http://www.anglistik.hhu.de/fileadmin/redaktion/Fakultaeten/Philosophische_Fakultaet/Anglistik_und_Amerikanistik/Ang3_Linguistics/Dateien/Detailseiten/Guenther/Guenther_2015_The_rich_the_poor_the_obvious.pdf)
- Hurst, E. (April 27, 2015). Heroic Maine governor Paul LePage stops poors from bogarting the baloney slices. *Wonkette*. Retrieved from <https://wonkette.com/583915/heroic-maine-governor-paul-lepage-stops-poors-from-bogarting-the-baloney-slices>
- Illegal, adj. and n. (2017) *OED Online*, Oxford University Press, Retrieved from [www.oed.com/view/Entry/91399](http://www.oed.com/view/Entry/91399).
- Kiritchenko, S. and Mohammad, S.M. (2016). Sentiment Composition of Words with Opposing Polarities. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, CA. Retrieved from <http://www.aclweb.org/anthology/N/N16/N16-1128.pdf>
- Kiritchenko, S., Mohammad, S.M., and Mohammad, S. (2016). Semeval-2016 Task 7: Determining Sentiment Intensity of English and Arabic Phrases. In *Proceedings of the International Workshop on Semantic Evaluation (SemEval '16)*. San Diego, California. Retrieved from [http://www.aclweb.org/old\\_anthology/S/S16/S16-1004.pdf](http://www.aclweb.org/old_anthology/S/S16/S16-1004.pdf)
- Landis, R. and Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *International Biometric Society*. 33 (1), 159-174. doi: 10.2307/2529310
- Lawler, J.M. (1973). Studies in English generics (Doctoral dissertation). University of Michigan, Ann Arbor, MI. Retrieved from <http://www-personal.umich.edu/~jlawler/dissertation/>
- Leichnitz, D. (2009, Oct.) I will worry about the gays when... [Web log post]. Retrieved from <https://blacknright.wordpress.com/chicagosgayhighschool/i-will-worry-about-the-gays-when/>

- Leichnitz, D. (2009, Oct.) Why I am "anti-gay" [Web log post]. Retrieved from <https://blackn-right.wordpress.com/chicagosgayhighschool/why-i-am-anti-gay/>
- Liu, B., Hu, M., and Cheng, J. (2005). Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th International World Wide Web Conference*. ACM.
- Mohammad, S., Dorr, B., and Dunne, C. (2009). Generating High-Coverage Semantic Orientation Lexicons from Overtly Marked Words and a Thesaurus. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*.
- Palmer, A., Robinson, M., and Phillips, K. (2017). *Illegal* is not a noun: Linguistic form for detection of pejorative nominalizations. In *Proceedings of The First Workshop on Detecting Abusive Language Online*.
- Pang, B., & Lee, L. (2018). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135. Retrieved from <http://www.cs.cornell.edu/home/llee/omsa/omsa.pdf>
- Poor, adj. and n. 1. (2017) *OED Online*, Oxford University Press, Retrieved from [www.oed.com/view/Entry/147749](http://www.oed.com/view/Entry/147749).
- Pullum, G.K. (1975) People deletion in English. *Working Papers in Linguistics*. Ohio State University 14, 95-101.
- Snow, R., O'Connor, B., Jurafsky, D., and Ng, A.Y., (2008). Cheap and fast---but is it good?: Evaluating non-expert annotations for natural language tasks. In *Proceedings of the EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing*.
- Stone, P. J., Dunphy, D. C., Smith, M. S., and Ogilvie, D.M. (1966). *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, MA: MIT Press.
- Sue, D. W. (2010). *Microaggressions in everyday life: Race, gender, and sexual orientation*. Hoboken, US: John Wiley & Sons, Incorporated.
- Sulis, E., Irazú, D., Farías, H., Rosso, P., Patti, V., and Ruffo, G. (2016). Figurative messages and affect in twitter: Differences between #irony, #sarcasm and #not. *Knowledge-Based Systems* 108, 132-143. Retrieved from <https://doi.org/10.1016/j.knosys.2016.05.035>
- van Dijk, T. A. (1993). Principles of critical discourse analysis. *Discourse & Society*. 4(2), 249-283. Retrieved from <http://journals.sagepub.com/doi/abs/10.1177/0957926593004002006>
- Warriner, A.B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45, 1191-1207. doi:10.3758/s13428-012-0314-x

- Wilson, T., Wiebe, J. and Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. *Proc. of HLT-EMNLP-2005*. Retrieved from <https://people.cs.pitt.edu/~wiebe/pubs/papers/emnlp05polarity.pdf>
- Wierzbicka, A. (1986). What's in a noun? (or: How do nouns differ in meaning from adjectives?). *Studies in Language*, 10(2). 353-389. 10.1075/sl.10.2.05wie